

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
Please do not report the images to the
Image Problem Mailbox.

(1)

(19) 日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表平6-508941

第6部門第3区分

(43) 公表日 平成6年(1994)10月6日

(51) Int.Cl.³
G 0 6 F 15/38
H 0 4 L 29/10

識別記号 庁内整理番号
Z 7323-5L
C 7323-5L
9371-5K

F I

H 0 4 L 13/00 3 0 9 Z

審査請求 未請求 予備審査請求 有 (全 17 頁)

(21) 出願番号 特願平4-503420
(86) (22) 出願日 平成3年(1991)12月30日
(85) 翻訳文提出日 平成4年(1992)8月31日
(86) 国際出願番号 PCT/US91/09761
(87) 国際公開番号 WO92/12494
(87) 国際公開日 平成4年(1992)7月23日
(31) 優先権主張番号 636, 400
(32) 優先日 1990年12月31日
(33) 優先権主張国 米国 (US)
(81) 指定国 EP(AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE), JP, US

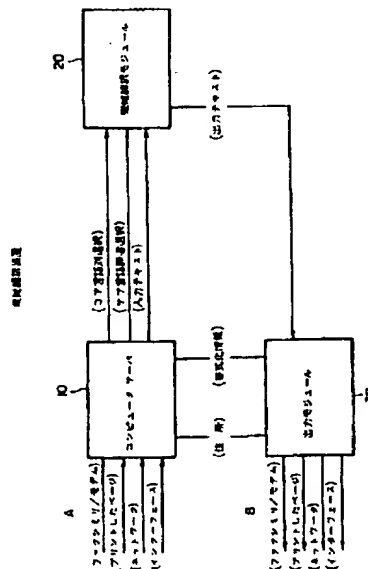
(71) 出願人 トランスリンク インターナショナル
コーポレーション
アメリカ合衆国 ニューヨーク州 10017
ニューヨーク 300 イースト 42nd
ストリート 18ス フロア
(72) 発明者 チョンレイトン ケイ
アメリカ合衆国 ニューヨーク州 10028
ニューヨーク 425 イースト 86ス
ストリート 7ス フロア
(72) 発明者 カンプラス クリスティーン ケイ
アメリカ合衆国 ニュブランズウィック州
68434 セワード 1556 ブレインビュ
ー

(74) 代理人 弁理士 田中 増顕

(54) 【発明の名称】 機械翻訳兼遠隔通信装置

(57) 【要約】

機械翻訳兼遠隔通信装置は、一般的な単語のためのコア言語辞書と、異なるドメインまたは使用者グループの特殊な単語のための複数のサブ言語辞書と、異なる使用者によって用いられる個々の単語のための複数の使用者辞書とを含む辞書データベース(22)を用いて、ソース言語の入力テキストをターゲット言語に自動的に翻訳する。装置は、電子テキスト、ファクシミリ(グラフ)入力、またはページ像データの形態の送信人からの入力を受信する受信インターフェース(11)と、任意の指定した受取人に翻訳した出力テキストを送信する出力モジュール(30)とを含む。入力テキストは、送信人、一人またはそれ以上の受取人、かれらの住所、テキストのソース/ターゲット言語、入力テキストに適用できる任意のサブ言語および出力テキストの任意の書式化要求事項を識別するカバーページまたは見出し(50)を伴う。装置はカバーページまたは見出しのデータを用いて、翻訳処理に使用されるコア言語、サブ言語および(または)使用者辞書を選択し、翻訳した出力テキストを書式化し、かつ指定した住所の受取人に出力を送信する。



請求の範囲

辞書データベース(22)は辞書保守ユーティリティの使用によってスクラッチエントリとしての新たな単語を使用者辞書に加え、スクラッチエントリをより高級なレベルのサブドメイン、ドメインまたはコア辞書を更新しおよび(または)移動させることによって時間の経過につれて新規なものを蓄積する。

1. 機械翻訳兼遠隔通信装置において、

(a) ソース言語の入力テキストからターゲット言語の出力テキストに機械翻訳を行う機械翻訳モジュールと、

(b) 翻訳された出力テキストが送信される受取人の住所を指定する少なくとも第1所定の領域を持つ制御部分を伴った翻訳すべき入力テキストを含む入力を第1遠隔通信リンクを介して受信する受信インターフェースと、

(c) 制御部分を電子的に走査し制御部分の第1の所定領域に指定された受取人の住所を認識するために受信インターフェースに接続された認識モジュールと、

(d) 機械翻訳モジュールによって翻訳された出力テキストを認識モジュールによって認識された受取人の住所に第2遠隔通信リンクを介して送信するための送信インターフェースを含む出力モジュールと、

を有することを特徴とする機械翻訳兼遠隔通信装置。

2. 請求の範囲第1項記載の機械翻訳兼遠隔通信装置において、受信インターフェースは光学ページスキャナ、ファクシミリモデムおよびネットワークから成る群のうちの1つまたはそれ以上の装置を通して入力を受信するように第1遠隔通信リンクに接続されていることを特徴とする機械翻訳兼遠隔通信装置。

3. 請求の範囲第1項記載の機械翻訳兼遠隔通信装置において、受信インターフェースへの入力はテキスト文字を符号化データであり、認識モジュールが像データを機械読取り可能なテキスト文字に変換する文字認識手段を有することを特徴とする機械翻訳兼遠隔通信装置。

4. 請求の範囲第1項記載の機械翻訳兼遠隔通信装置において、

機械翻訳モジュールは複数のソース/ターゲット言語にそれぞれ対応する複数のコア言語辞書を含み、制御部分は翻訳すべきテキストに適用可能な選択したコア言語辞書を指定する第2所定部分を有し、認識モジュールは制御部分の第2所定領域内の指定を認識して機械翻訳モジュールに選択したコア言語辞書を使用させる辞書制御手段を有することを特徴とする機械翻訳兼遠隔通信装置。

5. 請求の範囲第1項記載の機械翻訳兼遠隔通信装置において、機械翻訳モジュールは複数のサブ言語辞書を含む辞書データベースを有し、制御部分は翻訳すべきテキストに適用可能な第3の所定部分を有し、認識モジュールは制御部分の第2所定領域内のサブ言語指定を認識し機械翻訳モジュールに選択したサブ言語辞書を使用させる辞書制御手段を有することを特徴とする機械翻訳兼遠隔通信装置。

6. 請求の範囲第1項記載の機械翻訳兼遠隔通信装置において、送信インターフェースはファクシミリモデム、ネットワークインターフェース、およびプリンタから成る群のうちの1つまたはそれ以上の装置を通して出力を送信するために第2遠隔通信リンクに接続されていることを特徴とする機械翻訳兼遠隔通信装置。

7. 請求の範囲第1項記載の機械翻訳兼遠隔通信装置において、出力モジュールは制御部分の他の所定領域内で指定された選択したページ形式に機械翻訳モジュールからの翻訳された出力テキストを形式化するページ書式化モジュールを有することを特徴とする機械翻訳兼遠隔通信装置。

8. 請求の範囲第1項記載の機械翻訳兼遠隔通信装置において、受信インターフェースが第1遠隔通信リンクを介して使用者インターフェースを通ず使用者入力用のプログラムされた相互作用モジュールを有することを特徴とする機械翻訳兼遠隔通信装置。

9. 請求の範囲第8項記載の機械翻訳兼遠隔通信装置において、相互作用モジュールは使用者によって使用される制御部分のためにデータ記録を格納する使用者識別メモリに接続されており、データ記録は使用者によって選択され、翻訳すべき入力テキスト用の制御部分のために使用されることを特徴とする機械翻訳兼遠隔通信装置。

10. 請求の範囲第1項記載の機械翻訳兼遠隔通信装置において、機械翻訳モジュールは受信インターフェース、認識モジュールおよび出力モジュールから分離して翻訳を行う別個のプログラムされた処理手段を有することを特徴とする機械翻訳兼遠隔通信装置。

11. 機械翻訳装置において、

(a) 入力テキストと複数の可能なサブ言語から入力テキストに適用可能な選択したサブ言語を指定するサブ言語制御入力を受信する受信インターフェースと、

(b) ソース言語の単語に対応するターゲット言語の単語に対するエントリを含む辞書データベースを用いてソース言語の入力テキストのターゲット言語の出力テキストへの機械翻訳を行う機械翻訳モジュールと、

(c) ソースおよびターゲット言語の一般単語に対するエントリを含むコア言語辞書とサブ言語の特殊な単語のエントリを含む複数のサブ言語辞書を有する辞書データベースと、

(d) 入力テキストに適用可能な辞書データベースのサブ言語辞書を選択し機械翻訳モジュールに入力テキストの翻訳のための選択したサブ言語辞書を使用させるためにサブ言語制御入力に回答する辞書制御モジュールと、

(e) 機械翻訳モジュールによって翻訳されたターゲット言語のテキストを出力する出力モジュールと、

を有することを特徴とする機械翻訳装置。

12. 請求の範囲第11項記載の機械翻訳装置において、辞書データベースは複数のソース/ターゲット言語にそれぞれ対応する複数のコア言語辞書を有し、受信インターフェースは入力テキストに適用可能な選択されたコア言語辞書を指定するソース/ターゲット制御入力を受信し、ソース/ターゲット制御入力に回答する辞書制御モジュールが機械翻訳モジュールに入力テキストの翻訳のために選択したコア言語辞書を使用させることを特徴とする機械翻訳装置。

13. 請求の範囲第11項記載の機械翻訳装置において、受信インターフェースは入力テキストに関連する使用者を識別するデータを含むサブ言語制御入力を受信し、辞書制御モジュールは受信インターフェースによって受信した使用者識別データに従ってテキストに適用可能なサブ言語辞書を選択するための辞書選択モジュールを有することを特徴とする機械翻訳装置。

14. 請求の範囲第13記載の機械翻訳装置において、辞書選択モジュールは送信人および受取人の資格、性別、会社、地位、住所、使用者のグループおよび課題を含む使用者識別データから引き出される推論に基づいて適用可能なサブ言語辞書を選択する推論アルゴリズムを有することを特徴とする機械翻訳装置。

15. 機械翻訳装置において、

(a) ソース言語の単語に対応するターゲット言語の単語に対するエントリを含む辞書データベースを用いてソース言語の入力テキストのターゲット言語の出力テキストへの機械翻訳を行う機械翻訳モジュールと、

(b) ソースおよびターゲット言語の一般単語に対するエントリを含むコア言語辞書と使用者のグループによって用いられるサ

ブ言語の特殊な単語のエントリを含む複数のサブ言語辞書と、使用者の個別の単語用のエントリを含む複数の使用者辞書を有する辞書データベースと、

(c) 機械翻訳モジュールにコア言語辞書、任意の選択したサブ言語辞書、および使用者に属する入力テキストの翻訳のための任意の選択した使用者辞書を使用させるために機械翻訳装置に対する制御入力に回答する辞書制御モジュールと、

を有することを特徴とする機械翻訳装置。

18. 請求の範囲第15項記載の機械翻訳装置において、受信インターフェースはスクラッチ単語エントリを定義する使用者入力のためのプログラムされた相互作用モジュールを有し、辞書制御モジュールは対応する使用者辞書にスクラッチ単語エントリを加える使用者保守ユーティリティを有することを特徴とする機械翻訳装置。

17. 請求の範囲第18記載の機械翻訳装置において、辞書制御モジュールは他の使用者の辞書内の類似の単語エントリを参照して定義されるスクラッチ単語エントリを可能にする使用者保守ユーティリティを有することを特徴とする機械翻訳装置。

18. 請求の範囲第15項記載の機械翻訳装置において、辞書データベースは入れ子形式の構成を有し、使用者辞書が最も低級なレベルにあり、より高級なレベルのサブ言語辞書内に入れ子にされており、サブ言語辞書は最も高級なコア言語辞書に入れ子にされており、辞書制御モジュールは、機械翻訳モジュールによって、最も低級なレベルから最も高級なレベルの順に選択された使用者辞書、選択されたサブ言語辞書およびコア辞書にアクセスすることを特徴とする機械翻訳装置。

19. 請求の範囲第15記載の機械翻訳装置において、辞書制御モジ

明 細 書

機械翻訳装置通信装置

技術分野

本発明は、テキストの自動（機械）翻訳用装置に関し、特に、送信者からのテキストを他の言語に自動的に翻訳し、受取人に送信する遠隔通信に基づく装置に関するものである。

背景技術

数10年の開発の後に、人間の介在を最少にしてソース言語からターゲット言語へテキストを自動（機械）翻訳する分野は、商業に限られた、または言語環境に限られた機械翻訳装置で基本的レベルの許容できる翻訳テキストを作成できるという初歩レベルまで進歩してきた。いくつかの最近の装置は、選択した言語対、即ち選択したソース言語から選択したターゲット言語への制限のない入力に対する翻訳を作成することができ、翻訳は、ターゲット言語の自国作家に多分50%受け入れられるものである（任意の自盛り測定値を用いて）。翻訳装置に限られた適用領域（「サブ言語」という）の特定の語彙または統語スタイルに制限されると、達成できる結果は自国作家に受け入れられるレベル80%に近づくかもしれない。結果が大きく異なるのは、選択したサブ言語の限られた語彙および翻訳均等物と比較して、装置が語彙使用、統語形式、および表現において広範囲な変換を包含しなければならぬときの正確な翻訳を作成することの困難性によるものである。

特定のサブ言語適用に限定した機械翻訳装置の1つの例は、カナダ環境庁によって発表された天気予報を英語からフランス語へ翻訳するためにモントリオール大学によって開発されたTAUM-METEO装置である。TAUM-METEOは、(1) ソー

ジュールは辞書保守動作器(DMO)に単語エントリがドメインまたは使用者の使用者辞書に共通に表れるかを評価させ、そのような単語エントリをそのドメインまたは使用者のグループ用のより高級なレベルのサブ言語辞書に移動させる辞書保守ユーティリティを有することを特徴とする機械翻訳装置。

20. 請求の範囲第15項記載の機械翻訳装置において、辞書制御モジュールは単語エントリがドメインまたは使用者の使用者辞書に共通に表れるかを自動的に評価し、そのような単語エントリをそのドメインまたは使用者のグループ用のより高級なレベルのサブ言語辞書に移動させる辞書保守ユーティリティを有することを特徴とする機械翻訳装置。

ス言語の入力単語の順序および形態学的形式を分析してそれらの句および文を決定し、(2) 辞書検索と単語および句選択用の開発した転写規則を用いて入力テキストをターゲット言語の均等単語の文に転写し(直接的に翻訳し)、(3) ターゲット言語の統語および文法のための開発した規則を用いてターゲット言語の許容できる出力テキストに編集する。3つのステップから基本的に構成された翻訳の転写方法を用いている。TAUM-METOは気象学の少ないサブ言語(数百の地名を持った1,500辞書エントリ、テキストは時刻動詞を持たない)における英語-フランス語の翻訳用に動作するように設計された。このTAUM-METOは、入力単語の有効なレベルの形態学的分析の必要性を避ける、出力単語の選択および統語構造に対する選択の範囲を決めるドメイン特定単語のマーカのために入力テキストを分析し、出力単語および句の選択のために特有の転写規則を用いることによって、80%または90%の高レベルの翻訳正確さを得ることができる。

サブ言語翻訳装置の他の例は、遠隔通信の分野で、ドイツ語から英語のテキストへの大量の翻訳のためにオースチン所在のテキサス大学のリングイスティック リサーチ センターによって開発されたMETAL装置である。METAL装置も転写方法を用いるが、分析と転写の間に「集積」と呼ぶ第4のステップを加えている。集積ステップは、開発した文法および句構造規則に基づいて、一連の入力テキストの構成単語にテストを行い、それらの適用を制約することによって出力単語選択および統語の変化を減少しようと試みるものである。転写辞書は一般におよそ10,000単語対から成る。翻訳の品質の項目では、METAL装置は45%と85%の間で正確な翻訳を達成できたと報告されている。

ようとしている。さらに他のアプローチは、統語論的な中間構造に分解することと均等な出力句および文に転写することとを組み合わせている。機械翻訳の最近の開発のさらに他の議論に対しては、1987年ケンブリッジ大学プレスによって出版され、サージェイ ニレンバーグによって編集された機械翻訳、理論的および形態学的イシューおよび1990年8月オースチン所在のテキサス大学のリングイスティック リサーチ センターによって出版された「自然言語の機械翻訳における理論的および形態学的イシューに関する第8回国際会議の議事録」を参照されたい。

機械翻訳(MT)装置はより高いレベルの翻訳の正確さと有用性を提供するように開発されると予想される。しかしながら、基本の転写アプローチを用いる最近のMT技術は選択したサブ言語では許容できる翻訳の正確さを与えることができるが、未だ広範囲に使用されるには至っていない。MT装置の使用が限られている理由の1つは、最も最近の装置が単一の特定の用途、環境、言語対文脈に対して設計されているからである。文脈の要求事項が設計、文法の開発、辞書構造および品詞アルゴリズムの動機となっている。したがって、装置の有用性はその特定の文脈に制限されるようになる。このアプローチは、アプリケーションの範囲とそのようなアプリケーションおよび言語特定MT装置によってサービスを受ける使用者の範囲を著しく制限する。

発明の要約

したがって、本発明の主要な目的は、異なったソース言語、ターゲット言語およびサブ言語に対して機械翻訳を行い、翻訳したテキストを遠隔通信リンクを介して異なった言語および(または)の異なった位置の一人またはそれ以上の受取人に自動的に送信する装置を提供することにある。この装置は、許容できるレベルの

転写アプローチと統合する方法は中間言語アプローチであり、この中間言語は、ソース言語の入力テキストを「意味」または暗黙文脈を表す中間言語に分解し、次に、文脈的、語彙的、統語的規則の基礎知識を用いることによって、語彙構造をターゲット言語の均等な出力文に変換する。歴史的には、理解力の基礎を欠き独立の文の翻訳に制限された転写装置には、同音異義語、ポリセミック句および前文参照によって表される不明瞭さの面で正確な単語および句の選択を得る重要な問題があった。中間言語アプローチは、単一文より大きな文脈内のテキストの意味を表すことにより、理論的に、入力テキストの分析の不明瞭さを著しく減少させることができるので、好ましいものである。また、一旦入力テキストが語彙構造に分解されると、入力テキストは、各ターゲット言語用に開発された言語的および語彙的規則を用いて、多数のターゲット言語に理論的には翻訳可能である。しかしながら、実際には、中間言語アプローチは、ソース言語から中間言語に、さらにターゲット言語へ変換するための「意味」および理解力を表す世界的な記号言語の開発を必要とするので、実行が困難であることがわかった。中間言語装置の例としては、オランダのウトレヒトで行われている分布翻訳言語(DLT)およびカーネギーメロン大学の機械翻訳用のセンタの知識に基づく機械翻訳(KBMT)装置がある。

他の機械翻訳装置は、転写および中間言語アプローチの変形またはハイブリッドを用いて、開発されており、または開発中である。例えば、或る装置はテキストの不明瞭さを減少し、単語および句の選択の正確さを改良するために、人による予備編集および(または)後編集を使用する。他の装置は機械翻訳と向上のために基本の転写アプローチを知識および人工知能技術と組み合わせ

翻訳の正確さを提供できかつより高級なレベルの正確さおよび有用性に容易に高級化できるべきである。さらに他の目的は、そのような装置が人の介入が最少な状態で動作できると共に、新たな単語エントリを得てその辞書データベースに加えるための相互作用ユーティリティを有することである。そのような装置が、コア言語辞書、複数のサブ言語辞書および個々の使用者辞書を含む大規模な辞書データベースを、時間とともに蓄積し発展するような仕方、作成し構成することが望ましい。

本発明の基本的な態様に従って、機械翻訳装置遠隔通信装置は、

(a) ソース言語の入力テキストからターゲット言語の出力テキストに機械翻訳を行う機械翻訳モジュールと、

(b) 翻訳された出力テキストが送信される受取人の住所を指定する少なくとも第1所定の領域を持つ制御部分を伴った翻訳すべき入力テキストを含む入力を第1遠隔通信リンクを介して受信する受信インターフェースと、

(c) 制御部分を電子的に走査し制御部分の第1の所定領域に指定された受取人の住所を認識するために受信インターフェースに接続された認識モジュールと、

(d) 機械翻訳モジュールによって翻訳された出力テキストを認識モジュールによって認識された受取人の住所に第2遠隔通信リンクを介して送信するための送信インターフェースを含む出力モジュールと、

を有する。

サブ言語に関連する本発明のさらに特定の態様では、機械翻訳装置は、

(a) 入力テキストと複数の可能なサブ言語から入力テキストに適用可能な選択したサブ言語を指定するサブ言語制御入力を受

信する受信インターフェースと、

(b) ソース言語の単語に対応するターゲット言語の単語に対するエントリを含む辞書データベースを用いてソース言語の入力テキストのターゲット言語の出力テキストへの機械翻訳を行う機械翻訳モジュールと、

(c) ソースおよびターゲット言語の一般単語に対するエントリを含むコア言語辞書とサブ言語の特殊な単語のエントリを含む複数のサブ言語辞書を有する辞書データベースと、

(d) 入力テキストに適用可能な辞書データベースのサブ言語辞書を選択し機械翻訳モジュールに入力テキストの翻訳のための選択したサブ言語辞書を使用させるためにサブ言語制御入力に应答する辞書制御モジュールと、

(e) 機械翻訳モジュールによって翻訳されたターゲット言語のテキストを出力する出力モジュールと、

を有する。

本発明では、サブ言語制御入力は、入力テキストのより正確な翻訳を行うために、用いられる入力テキストに適用できると思われる選択されたサブ言語辞書を生じさせる。辞書データベースは、異なるソース/ターゲット言語およびサブ言語用のコアおよびサブ言語辞書を含む。この多数のコア言語およびサブ言語能力を持つ機械翻訳装置は送信人からのテキストを翻訳して他の言語の一人またはそれ以上の受取人に送信する遠隔通信装置に用いられる。入力テキストに伴われるカバーページまたは見出しは、選択したソース/ターゲット言語、適用できるサブ言語、受取人の住所(電子、ファクシミリ、郵便)を指定するのに用いられる。

好ましい実施例では、受信インターフェースは通信ラインを通しての電子(機械読取り可能な)テキストまたはファクシミリ/

モデムボードスキャナを介したページデータのような入力テキストを受信する。受信インターフェースは任意のページ像データを電子テキストに変換する認識モジュールと共にコンピュータサーバ内で動作させられる。認識モジュールは入力テキストに適用できるソース/ターゲット言語およびサブ言語の選択を決定するために入力テキストに伴われたカバーページまたは見出しの指定を定直し認識する。電子テキストの場合には、カバーページおよび入力テキストは、ディスクファイルによって、電子ファイルをダウンロードすることによって、またはオンラインの使用装置相互作用によって、導入されてもよい。適当なカバーページの発生を容易にするために、オプションの相互作用モードによって使用者識別、サブ言語、選択に関する情報を使用者に勧める。推論アルゴリズムが使用者およびカバーページ情報を評価して適用可能なサブ言語辞書を決定するのに用いられてもよい。

出力モジュールは翻訳された出力テキストを特定の受取人またはターゲット言語に適した所望のページ書式(フォーマット)に編集するページ書式化プログラムを有してもよい。出力モジュールは、また、ターゲット言語の代用翻訳と共に(または)単独に、オリジナルソース言語の入力テキストの不明瞭な句の脚注を与え、脚注機能も有してもよい。出力モジュールは、ファクシミリ送信のためにファクシミリ/モデムボードに接続された送信インターフェース、出力ページを印刷するプリンタ、出力電子テキストを受取人の電子住所に送信するための遠隔通信インターフェースを有する。受信インターフェース、辞書データベース、辞書制御モジュール、機械翻訳モジュールからの出力モジュールの調整によって、機械翻訳の改良が進むにつれて、機械翻訳モジュールは、装置の他の部分を調整させるまたは時代遅れのものにさせない

で、高級化され、または置き換えられる。

機械辞書データベースに関連する本発明の他の態様では、機械翻訳装置は、

(a) ソース言語の単語に対応するターゲット言語の単語に対するエントリを含む辞書データベースを用いてソース言語の入力テキストのターゲット言語の出力テキストへの機械翻訳を行う機械翻訳モジュールと、

(b) ソースおよびターゲット言語の一般単語に対するエントリを含むコア言語辞書と使用者のグループによって用いられるサブ言語の特殊な単語のエントリを含む複数のサブ言語辞書と、使用者の個別の単語用のエントリを含む複数の使用者辞書を有する辞書データベースと、

(c) 機械翻訳モジュールにコア言語辞書、任意の選択したサブ言語辞書、および使用者に属する入力テキストの翻訳のための任意の選択した使用者辞書を使用させるために機械翻訳装置に対する制御入力に应答する辞書制御モジュールと、

を有する。

本発明では、大容量の辞書データベースが保守され、このデータベースは用法の異なる階層レベルにおいて言語学的に特定された単語エントリを含む辞書を有する。最も下級の(使用者の)レベルにおいて、特定の使用者は一時的な単語または「スクラッチ」単語エントリをそれぞれの使用者辞書に入れることができる。機械翻訳装置は、使用者によって特に使用される特有のまたは新たな単語または句を含んでもよいテキストの機械翻訳を行うために特定の使用者辞書を使用する。辞書制御モジュールは辞書保守ユーティリティを含み、このユーティリティによって、使用者がそのようなスクラッチエントリを使用者の辞書に入れることを可

能にし、スクラッチエントリを機械翻訳のために有効な辞書エントリとして確認できるように辞書保守動作器(DMO)がスクラッチエントリを見直すことを支援する。辞書保守ユーティリティは、より低レベルの辞書に現れた単語エントリをより高レベルの辞書に移動すべきであるか否かを評価するための自動化したプログラム化した手順を含む。

本発明の他の目的、特徴および利点は、添付図面を考慮して本発明の好ましい実施例の以下の詳細な記載から明かになる。

図面の簡単な説明

第1図は、本発明の機械翻訳装置遠隔通信装置の概略図である。

第1A図は、受信インターフェース、認識モジュール、辞書制御モジュールを含み、機械翻訳モジュールおよび出力モジュールに接続されたコンピュータサーバの概略図である。

第1B図は、翻訳処理モジュール、辞書データベース、およびコンピュータサーバと出力モジュールへの接続を含む機械翻訳モジュールの概略図である。

第1C図は、ページフォーマット化モジュールおよび送信インターフェースを含む出力モジュールの概略図である。

第2図は、コア言語辞、サブ言語、受取人情報および添付テキストページを指示するカバーページの図である。

第3図は、入力表意文字テキスト、およびページフォーマット化機能を用いて機械翻訳装置によって実行された出力英文の図である。

第4図は、辞書選択部、および(相互作用)使用者保守モジュールと辞書保守モジュールを含む保守サブモジュールを含む辞書制御モジュールの概略図である。

第5図は、装置の使用者との相互作用のための相互作用入力端

集替の概略図である。

第6図はサブ単語軸からスーパー単語軸へのエントリを抑制および推通するための辞書保守ユーティリティを示す概略図である。

第7A図は、辞書保守ユーティリティの機能として、同一単語エントリからスクラッチ単語エントリの生成を示す。

第7B図は、スクラッチ単語エントリに基づく単語エントリのための辞書階層の種々のレベルを走査するために相互作用人力編集器と共にユーティリティを用いることを示す。

第7C図は、スクラッチ単語エントリが生成される同一単語エントリの代表的な内容を示す。

第7D図は、同義語の単語エントリから「コピーキャット」単語エントリの生成を示す。

好ましい実施例の詳細な説明

第1図を参照すると、好ましい態様の本発明の機械翻訳装置は、コンピュータサーバ10、機械翻訳モジュール20および出力モジュール30から成る。(装置のこれらのコンポーネントおよび他のコンポーネントは明瞭にするために大文字を付している。) コンピュータサーバ10は、遠隔通信リンクAとして一般的に表示された複数の入力ソースのうちの任意のものからカバーページ即ち見出し部を伴った電子テキスト入力を受信する。コンピュータサーバ10は、コア言語および翻訳すべき入力テキストに適用できるサブ言語を指示するカバーページ即ち見出し部内の制御データを認識する機能を持つ。コンピュータサーバは、また、遠隔通信リンクBとして一般的に指示された複数の出力装置のうちの任意のものを介して指定された受取人に翻訳されたテキストを送信するために出力モジュール30によって用いられる出力アドレスおよびページフォーマット化データを認識する。装

ページ指示または使用者指示参照カバーページデータを認識する。辞書制御モジュール13は辞書選択モジュールを含み、この辞書選択モジュールは認識モジュール12から受信した制御データを評価し、機械翻訳モジュール20によって用いられる適切なコア言語およびサブ言語辞書を指示する。辞書制御モジュールは、また、辞書保守モジュールを含み、この辞書保守モジュールによって辞書保守オペレータ(DMO)が辞書データベース22に辞書エントリを生成しかつ更新することが可能になる。

入力テキストを伴うカバーページまたは見出しからの制御データを用いて、コンピュータサーバ10により装置は、入力テキストのソース言語の送信者の指定、出力テキストのターゲット言語、特定のドメイン内で用いられる任意の特定のサブ言語、使用者群または通信形式、出力テキスト用の任意の好ましいページフォーマット、および出力が送られる一人またはそれ以上の受取人のアドレスを自動的に認識する。したがって、装置は、異なるソース/ターゲット言語およびサブ言語用の辞書データベース22に維持された指定のコアおよびサブ言語辞書に自動的にアクセスでき、実質的に人が介入する必要性なしに、翻訳したテキストをフォーマット化して遠隔通信リンクBを介してそれぞれのターゲット言語で受取人に送信する。

コンピュータサーバ10は複数の受信装置と接続している。例えば、入力データはサーバ装置用の1/0バスに差し込まれたファクシミリ/モデムボードを介してファクシミリ伝送として受信できる。このようなファクシミリ/モデムボードは広範に使用されており、サーバ装置内のそれらの動作はこの技術分野の当業者にとって十分わかっている。入力、電話線に送信された信号から変換されたファクシミリページを印刷する、電話線に接続され

るのモジュール化により、機械翻訳モジュール20は、装置の他の機能を不良化または陳腐化させることなく、オペレータ保守によって更新され、高級化され、または置換されてもよい。

機械翻訳モジュール20はソース言語の入力テキストからターゲット言語の出力言語に機械翻訳を行うことができる。本明細書で述べる機械翻訳(MT)装置の例として、ソース/ターゲット単語エントリの検索のための機械読取り可能辞書の使用に基づく転写(トランスファ)形式のMT装置を参照されたい。本発明の原理は、中間言語形式のMT装置にも適用されてもよい。転写形式のMT装置は、中間言語装置よりも代用用語に対して広く受け入れられており、ソース/ターゲット言語対用に設計された機械辞書に組み入れられた言語知識に主に依存している。転写形式のMT装置の動作は機械翻訳分野の当業者には十分理解されるので、本明細書ではこれ以上説明しない。

入力データの受信および抽出

第1A図は、遠隔通信リンクAに接続された受信インターフェース11、認識モジュール12および辞書制御モジュール13を持つコンピュータサーバ10を示す。受信インターフェース11は相互作用モードプログラム(後述する)を含んでもよく、この相互作用プログラムによって、使用者はカバーページまたは見出し部を与えることができ、使用者の通信に関連する翻訳パラメータに属する使用者IDファイルを更新できまたは生成でき、または相互作用テキストエントリセッション中特殊な使用者の辞書エントリを生成することができる。認識モジュール12は、文字認識(しばしば「OCR」という)プログラムを含み、この文字認識プログラムは、ページイメージデータを認識して機械読取り可能テキストに変換し、かつ使用者IDファイルに格納されたカバ

た従来のファクシミリ機械から受信されることもできる。紙送り器を持った従来のページスキャナをコンピュータサーバの入力用のページイメージデータとしてのファクシミリまたは印刷されたページを走査するのに用いることができる。ページイメージデータは次にOCRプログラムによって機械読取り可能なフォームに変換される。入力は電子テキストまたはテキストファイル(ASCIIのような)としての遠隔通信プログラムまたはネットワークインターフェースを通して受信されてもよく、この場合、OCRプログラムによる変換は必要ない。

OCRプログラムは入力データの受信を取り扱うインターフェースプログラムと共にコンピュータサーバ10内のアプリケーションプログラムとして存在してもよい。OCRプログラムは広範囲に使用されており、それらの動作はこの技術分野で周知である。例えば、日本語のカナおよび表意文字を認識するOCRプログラムは、日本国東京所在のカテナコーポレーション(Catena Corp.)によって提供されている。英数字文字用のOCRプログラムの例は、カルフォルニア州サンタクララ所在のカレラレコグニションシステムズ(Calera Recognition Systems)によって提供されたワードスキャンTM(WordScanTM)である。コンピュータサーバ10は好ましくは高速マルチタスクングPCコンピュータまたはワークステーションである。

第2図を参照すると、コンピュータサーバ10は、カバーページ、即ち見出し50と入力テキスト60の2つの部分に分割される入力データを受信する。図示の例では、カバーページはページ配向装置内の入力テキストの他のページに関連して用いられる。電子テキストファイルまたはテキストメッセージの伝送の場合、

通信用の前の見出しまたは識別子が用いられる。カバーページ50は、テキストのためのソース/ターゲット言語、サブ言語、ページページフォーマットおよび受取人の選択を指示する多数のフィールドを持つ。カバーページ50は、所定のフォーマット内のデータフィールドで構成され、所定のフォーマットは、所定のフィールド内の制御データを容易に認識できるように、コンピュータ10の認識モジュール12によって容易に認識される。

例えば、カバーページ50は、高レベルの信頼性を持った光学的走査のために、印刷したページ上のフィールド境界線とマーキングでレイアウトされフォーマット化されてもよい。分割線51および大きな形式のサイズ見出し52が、送信者、ソース/ターゲット言語、サブ言語（通信形式または課題）、ページフォーマットおよび受取人のアドレスフィールドにマークを付けるために使用されてもよい。ボックス53（マークを付けられ、またはその内部を黒く塗りつぶされる）によって、指示された選択が誤りなく決定される。送信人および受取人の名前、それぞれの会社名、住所（アドレス）、および電話および（または）ファクシミリ伝送番号は、それぞれのフィールド51、52が区別されると、文字認識によって決定される。任意のページ長の入力テキスト80がカバーページ50に続く。他の方法としては、カバーページまたは見出しによって通常供給される情報は、使用者のIDファイルに格納されて、使用者の選択に応じて記憶された筆記として自動的に供給される。

カバーページまたは見出しのフィールドから許容選択に関連するデータを抽出するのは、認識モジュール12の仕事である。バッチモードでは、このデータは予め決定されている。データは、使用者によって各特定の翻訳トランザクションでカバーページフ

ィールドに 含まれるか、または参照符号によって認識モジュール12にある使用者識別（ID）ファイルに供給される。

受 インターフェース11を介してカバーページまたは見出しを特定化する相互作用モードにおいて、最初に使用者に所定のセットの書込みデータが与えられ、次に別の値が勧められるか、またはいろいろな別の値が与えられる。別の値は、使用者IDファイルに既に格納されたデータに基づくか、または使用者によって入力されたデータから引き出される推論に基づくものである。例えば、使用者Aは受取人Zを名前だけで特定してもよく、次に、確認または訂正のために、受取人Zの住所、資格、関連会社のような付加的データが与えられる。他の方法としては、受取人Zは過去に使用者Aによってはアドレスが与えられていないが、ドメイン（このドメインは使用者Aも一員であるドメインである）中に分類される使用者である場合には、用いるか否かは任意であるが、ドメインのサブ言語群が使用者Aに与えられるという推論が行われる。

相互作用モードでは、通信形式、課題ドメイン、ソーシャルインディケータ等を含む所定の翻訳のために1つのまたはそれ以上のサブ言語を選択するのを援助する値を認識させるかまたはフィールドから選択するのを使用者に勧める。カバーページ情報の書込みを自動化することによって、装置は、供給されるデータの完全性と結合性を制御し監視しながら、使用者のためにコンピュータ化した能力を用いることができる。

カバーページは複数のアドレス位置およびターゲット言語の複数の受取人を指定してもよく、アドレス位置およびターゲット言語の各々は出力のために特定のフォーマット化の要件を持ってもよい。自動化した支援により、将来の受取人の各々に使用者ID

ファイルに格納されたデータに割り当てられる識別コードを付すことができる。例えば、旅行業者は、色々な通信モードでアクセスでき、広告物を普通送る色々な場所や言語の決まった顧客を持っている。一人の顧客は、印刷して普通郵便で送ってもらうために、縦書きの表紙文字を「右から左」に並べるようにフォーマット化した日本語の翻訳を必要とするかもしれない。他の顧客はファクシミリで送られる独語の翻訳を必要とするかもしれない。さらに他の顧客は電子メール（E-mail）の必要性があり、また印刷したコピーも必要としているかもしれない。住所や要求事項のこれらの組み合わせが予め決められて使用者のIDファイルに格納される。これらの住所の各々に対するカバーページフィールド用のデータを住所のアルファベット名のような簡略コードに割り当てることができ、また認識モジュールによって使用者IDファイルから取り出すことができる。

使用者IDファイルは機械翻訳サービスへの使用者による予約時に確立されてもよく、その後ととき更新される。相互作用モードを用いて、使用者には、名前、性別、資格、会社、住所、グループ関連会社、ソース言語等ばかりでなく、カバーページに自動的に書き込むために使用者IDファイルに格納された将来の受取人または受取人のグループに関連するデータを供給することが勧められてもよい。使用者に適切なサブ言語選択は使用者の要求事項とサービスに予約する他の使用者の要求事項を比べることによって識別されるか、または質問されてもよい。

使用者には、翻訳の提出に期待される代表的なテキストのサンプルばかりでなく用語の同意語反意語辞典のための個々の単語や重要単語を与えることが勧められてもよい。自動ユーティリティが、同一の課題ドメイン内の個々の使用者および使用者のグルー

プによって供給された関連用語の同意語反意語辞典に部分的に基づいて使用者に最も適用可能なサブ言語群を決定するために用いられてもよい。1つまたはそれ以上の確立されたグループまたは課題のドメイン内の使用者であることを認識することは許容選択と保守の重要な部分である。

各翻訳トランザクションのときに、使用者には、カバーページを開始すべき特殊な翻訳トランザクションに合わせるように装置によって勧められてもよい。装置は使用者に欠陥のあるカバーページの構成を認識させ、予め確立したカバーページ構成を選択させるかまたは変更させ、または白紙のスクラッチカバーページまたは使用者IDファイルからのデータで部分的に書き込まれたスクラッチカバーページに書き込ませる。

サブ言語群を用いる機械翻訳

第1B図に示すように、機械翻訳モジュール20は、翻訳処理モジュール21と多数のコアソース/ターゲット言語I、II、III、IV等用の辞書を含む辞書データベース22から成り、コアソース/ターゲット言語の各々は複数のドメイン、サブドメインおよび使用者辞書を含んでもよい。翻訳処理モジュール21は、ユタ州プロボ所在のエグゼクティブ コミュニケーション システムズ インコーポレーテッドによって提供されているECS ナチュラル ランゲージ プロセッシング システム（以後「ECS/MTシステム」という）のような従来の転写形式のシステムであってもよい。辞書選択モジュールによって与えられるコア言語およびサブ言語の選択インデックスは辞書データベース内のどの辞書を用いるかを決定する。選択された辞書は、動作辞書と共にコンパイルされるか、またはシステムRAMメモリ内に優先されて階層的に配列される。

機械翻訳モジュールは、用いられるコア言語と（もしあるならば）サブ言語、および辞書制御モジュール18を介してコンピュータサーバ10からの入力テキストを指定するシステムデータを受信する別個のモジュールとして図示されている。このように、翻訳機能は、受、編纂、使用者ID、辞書選択、辞書保守およびコンピュータサーバ10のその他の機能とは別に維持されており、このため、翻訳機能はシステムの残りの部分に対して中断させることなく容易に更新され、および（または）向上したプログラムに置き換えられる。コンピュータサーバ10は、認識モジュール12によって抽出されたデータに基づいたコア言語およびサブ言語を選択する辞書制御モジュール13の動作によって機械翻訳モジュール20に対する制御ユニットとして働く。

コンピュータサーバ、機械翻訳モジュールおよび出力モジュールのすべては同一のワークステーションに一緒に存在する。機械翻訳装置に対する最近の目標は約20,000〜30,000語/時間の速度である。最近入手できる転写形式の翻訳プログラムを用いるワークステーションは約50〜100MIP（1秒当たり百万指令）の処理速度でこの範囲を達成できる。ディスクアクセス時間の節約が、選択したコア言語およびサブ言語辞書のすべてをアクティブメモリに保持するのに十分なRAM容量を与えることによって得られる。ソース、転写、およびターゲット語彙用の80Kエントリ（各100バイト）の代表的なコア辞書の寸法各々プラス5K×3エントリの3つのサブ言語辞書、さらにシステムプログラムおよびオペレーションファイルに対して、内部メモリの48MB程度またはそれ以上のRAM容量が望ましい。

他の方法としては、システムには別個の処理ユニットが設けられてもよい。例えば、コンピュータサーバおよび出力モジュール

が遠隔通信ワークステーションとして設けられてもよく、一方機械翻訳モジュールが多数のソース/ターゲット言語、サブ言語および出力フォーマットの高速パッチ処理のためのRISCプロセッサ、並列プロセッサまたはスーパーコンピュータを介して実行されてもよい。

機械翻訳は、一般に、翻訳すべきテキストの各文を一連の段階を通過させることによって行われる。通常は、これらの段階には、(a) ソーステキスト辞書の検索および形態学的分析、(b) 表意文字の識別、(c) 複合名詞の識別、(d) 前置詞の処理、(e) 名詞および動詞句の識別、(f) 主語-述語識別、(g) 統語論多義性識別、(h) イディオムの処理、(i) ソース構造をターゲット構造に配列すること、(j) ターゲットテキストの統合および形態学的処理、および(k) ターゲットテキスト内の単語および句の後配列がある。

例として、ECS/MTシステムは辞彙機能文法理論および構造に基づいた転写システムである。入力文は左から右に1単語ずつ分析される。各単語は、ソース辞書の調べて調査されてその形態学的、語彙的および統合論的属性が決定される。辞彙機能文法のECS実行では、単語の指示された属性が分析ルーチンを呼出したまたは文法の規則を呼出すのに用いられ、文法の規則は文の句成分内の単語の位置および機能の認識を可能にする。分析規則および分析処理に基づく決定がソース言語の入力文の語彙の意味と句の構造を明確にする支援となる。

分析段階での分析の結果は、文のソース言語の句構造を表現する中間的なグラフまたはテーブルであり、このグラフは文および語彙の属性内で単語の文法上の機能を表示する直接式非線形のグラフに図示される。この辞彙機能文法表現はかなり言語的に独立

性がある。転写段階中、ソース言語文の機能構造表現は辞彙および統語的転写規則によって機能構造および語彙属性の均等なターゲット言語表現に転写される。次に、このターゲット言語表現はターゲット言語辞書によって与えられる語彙データおよび文法規則を用いて出力文に合成される。ソースおよびターゲット単語エントリ、バイリンガル転写エントリ、およびソースおよびターゲット言語の両方に対する形態学的、統語的および語彙的規則を含むコア言語辞書が各言語対に対して必要とされる。

本発明では、翻訳処理モジュール21は、特定化した単語エントリと、入力テキストに特に適用できるサブ言語に特定した文法規則を含む選択したサブ言語辞書を用いる。辞書データベース22に設定されている各サブ言語辞書は、取扱い可能な寸法、表現の予想可能なモード、および統語的構造を持ち、さらに同音異義語、多義性句および特定化した参照事項に対する十分理解できる文脈を持つように選ばれる。

機械翻訳の分野では、用語「サブ言語(sublanguage)」は、普通、限定した組の用語を持つ認識したドメインとそのドメインを特徴づける言語用法のパターンを意味する。本発明では、この「サブ言語」は、任意の組の用語および用法の分野、使用者のグループさらに個々の使用者に属する用法のパターンを意味するように一層広義に用いられている。即ち、「サブ言語辞書」は、好ましい組の用語および用法が識別できる限り、変更可能なまたは「特別の」基礎の上に設定される。

第2図に示すように、例えば、指定されたサブ言語は、ビジネスレター、法律/技術分析、技術文書、経費/市場報告書、または一般文書のような通信形式を含む。ビジネス通信は、通常、数ページであり、限られた用語数(5000単語程度)と限られた

組の統語的構造（多くは、平叙文に限られる）を用いる。

指定したサブ言語は、また、特定の分野、例えば、物理、化学、電子、軍事等のような技術分野、または、旅行、不動産、金融、出荷、保険等の市場分野、または同業者、会社、デパートまたは互いに普通の通信関係にある単なる個人のような使用者のグループを包含してもよい。

サブ言語辞書は、社会的に決められた用法や特定の文脈に対応するように設定されてもよい。例えば、日本語のような成る言語では、幾つかの単語、住所のフォームおよび全体の文章が、歳上か歳下か、性別、地位、グループ（内部の人/外部の人）、話し手の事情、宛て先の人によって決定される。このような特定の用語および用法は、通信のためのカバーページまたは見出しで識別されるファクタ、例えば送信者および受取人の資格または性別指示表題、それぞれの会社における地位、送信者および受取人の住所に従ってアクセスできる特有のサブ言語辞書として設定される。

サブ言語辞書の設定は機械翻訳装置に最近用いられている辞書強化工法で実行されることが出来る。例えば、ECS/MT装置は選択したサブ言語用の用語を含む選択した言語対用の辞書を開発するための一組の工具を提供する。規則編集工具によって言語学者はサブ言語のために形態学的規則、句構造の規則および転写規則を制作し変更できる。辞書保守ユーティリティによって、辞書への、ソースエントリ、ターゲットエントリ、ソースからターゲットへの転写エントリを含む語彙エントリの創作および変更が可能である。翻訳モジュールは、辞書に編集された、言語学的なテーブル、分析規則、転写規則および語彙選択エントリを用いるテーブル駆動式翻訳を行う。形態構造モジュールは規則を適用し

て形態学的な複合語を分析し、ソース語彙用語の辞書検索のために辞尾変化してない形状を決定し、ターゲット言語の形態学的複合語を発生させる。辞書選択成分は、好ましい辞書関係に基づいて辞書事項に対する語義属性の割当て、好ましい翻訳の語彙的明瞭さと選択に対するこれらの属性の接近性と整合性を行う。

辞書組織および選択

本発明では、コア言語辞書および複数の固有のサブ言語辞書（形態学的および統語論的分析および発生のための語彙エントリと文法規則の両方を含む）が装置の辞書データベース22に維持されている。コア言語辞書は、機械翻訳の分野で普通に使われる言語学的方法と工具に従って開発され維持される。本発明では、サブ言語は、任意の識別された商業上のまたは技術上の分野、アプリケーションドメイン、使用者のグループおよび個々の使用者に対して設定される。サブ言語の境界線や一般のサブ言語のパターンを厳密に識別する特別な努力はなされない。その代わり、サブ言語は、一人の使用者またはグループ内の使用者の語彙または統語選択が識別されるならばいつでも、設定されまたは更新される。個々の使用者または低級レベルの辞書が、一層一般的なサブ言語の語彙または統語選択が識別されるときには、任意の分野、アプリケーションドメインまたは使用者グループ用のマスター（高級レベルの）サブ言語に組み合わされるかまたは組み込まれる。

辞書データベースは、直ちに、エントリ内の言語の属性の用法および繋がり、構造上の関係の一般法則の順に配列した入れ子にされた辞書内の階層構造を実施し、それによって、種々の辞書とこれらの辞書内の特定のエントリが課題ドメインとこれらの課題ドメインに属するサブ言語の辞書の確立を知らせる。辞書データ

ベースはコア言語辞書を含み、このコア言語辞書は最も一般的な用法の単語に対するエントリばかりでなくそれらの形態学的および統語的構造を分析し発生させるための一組の文法規則を含む。転写アプローチでは、コア辞書は、翻訳すべき各用語に対して3つの並列なエントリを含み、即ち、それ自身の言語に関連する単語の形態学的、統語的および語義的特性に関する情報を含む2つのモノリンガルエントリ（1つはソース言語に対するものおよび他の1つはターゲット言語に対するもの）と、翻訳中どのような構造の変更がなされるかに関する情報を含むソース単語をターゲット単語に翻訳するのに必要である詳細を特定するバイリンガル転写エントリとを含む。モノリンガルエントリは他のソース/ターゲット言語対に対するモノリンガル辞書で使用可能であってもよいが、バイリンガル（転写）エントリは関連する言語対に対して特定なものである。

第1図に示すように、本発明の辞書データベースは、コア言語辞書と共に、入れ子にされたサブ言語辞書の多数のレベルを考慮する。最も低級のレベルでは、使用者は個々の使用者用に存在してもよい。使用者辞書はより高級なレベルの「サブドメイン」または「マスク」辞書Sub1、Sub2等内に入れ子にされる。サブドメイン辞書はより一般的な単語エントリと文法/言語的規則を含み、これらはそのサブドメイン内にグループ分けされた使用者に共通であるかまたはサブドメインに引照される。サブドメイン辞書はより高級な「ドメイン」辞書Dom1、Dom2等内に入れ子にされている。ドメイン辞書はより一般的な単語エントリと文法/言語的規則を含み、これらはそのドメイン内にグループ分けされた使用者に共通であるかまたはドメインに引照される。最も高級なレベルでは、ドメイン辞書はコア辞書内に入れ子にさ

れ、コア辞書は共通単語および規則を含み、これらは含まれるドメインの大部分または全てに対して包括的である。サブ言語辞書エントリはコア辞書エントリと同一の一般的な構造を有する。このように、転写アプローチでは、各入力単語用の2つのモノリンガルおよび第3の転写エントリはサブ言語辞書で利用できる。

辞書のこの階層組織によってサブ言語-特定テキストの翻訳で最少辞書検索時間が得られる。その理由は、最初に使用者辞書に対する検索に向けられ、次にサブ言語辞書に向けられるからである。より低級レベルの辞書を最初に検索すると、広範囲のより一般的なコア辞書が与える以上の一層正確な、効率的な、慣用的な翻訳が得られる。もしより低級レベルの辞書が人力用語を分析して回答を与えることができないならば、辞書制御モジュール13が次のレベルの辞書にアクセスし、より特定のレベルでの翻訳に失敗すると最終的にはコア辞書にアクセスする。辞書は辞書選択モジュールによってドメインの特性に対して選択されるが、アクセスの進行は入れ子構造に特有のものである。

辞書データベースの他の特徴は、言語情報の処理にある。人の文法の発生は、言語データを、名詞、動詞、形容詞等の類型に編集し、他動詞、自動詞等の至類型に編集し、即ち、類型の階層に編集することができることによる。このデータに基づいて分類をコンピュータ的に利用できる。本発明の好ましい実施例では、言語データは、周知の対象物指向プログラミング技術を用いて、対象物として特定されてもよい。この技術によってメッセージが対象物に属するデータの操作に関して対象物間を通ることができ、対象物は、文節である対象物の広がった分類に対して一般的な特徴を引き継ぐ。サブ言語辞書組織によって特定の特性と処理を同一の単語に対する任意のコア辞書のエントリとは独立にサブ言

語辞書内のエントリに対して割当てることができ、このため、ドメインで特定された単語のサブ言語使用が可能となる。階層の数レベルにおけるエントリに共通な言語の特徴と処理がすべてのレベルで使用でき、言語分析の能率的な保守が可能である。

文法的に関係ある語（前置詞、接続詞等）の面では、辞書データベースは階層内の並列なレベル上の辞書にアクセスでき、比較できる。数人の使用者の使用者辞書からの同一ドメインまたはグループ内への同一のエントリは、ある信頼性の基準の適用と履行の際に使用者辞書から出され、より高級なレベルの辞書に「昇格」されてもよい。辞書保守動作部(DMO)に利用できるユーティリティ（後述する）が、オーバーラップしたまたは矛盾するエントリを検出して訂正するために、使用者、サブドメインおよびドメイン辞書内のエントリ間の関係を調査することができる。

第4図に示すように、辞書制御モジュール13は2つの重要な機能、即ち、辞書選択および辞書保守を有する。その辞書選択能力13aでは、認識モジュール12によって与えられた制御データに基づいて、機械翻訳モジュール20によって用いられるようにコアおよびサブ言語辞書の選択を決定する。辞書選択機能は、認識モジュール12からのカバーページまたは見出しにアクセスでき、機械翻訳モジュール20に供給される辞書選択を決定する階層機能を含む。

辞書選択は翻訳処理のときに利用される。この辞書選択は、与えられたテキストの翻訳のために用いられる適切なコアおよびサブ言語辞書を決定するために、認識モジュール12から受信したカバーページまたは見出しデータに選択アルゴリズムを適用する。例えば、データはコア言語対、課題ドメイン、通信形式および社会的表示（「Mr.」または「Mrs.」、敬称等）に関する情報

をもち、これらのすべては、適用できるサブ言語辞書を選択するために推論機関のアルゴリズムによって使用されてもよい。各サブ言語辞書が設定されると、用法パラメータに関する仕様が辞書制御モジュール13内で割当られる。

サブ言語辞書選択と順序アルゴリズムに対する別のアプローチが、カバーページから得られるデータの類似および装置全体の他の態様から用いられてもよい。例えば、最も簡単なレベルでは、辞書制御モジュール13の選択アルゴリズムは、カバーページまたは見出しの表数、例えば、送信者の名前、「通信形式」、課題表示「re」等に直接対応するサブ言語を指定できる。

辞書保守

辞書保守モジュール13bは、多数のコア、サブ言語および使用者辞書に単語エントリーを入れ、書き込み、保守する。辞書保守モジュール13bは、使用者辞書内に一時的「スクラッチ」エントリーを作成するための入力編集器を持った相互作用使用者保守能力部13b-1と、より一般的なDMO支援ユーティリティによって分析されて供給されたデータに基づく辞書エントリーを更新するためのプログラムされた辞書保守ユーティリティを持ったDMO保守能力部13b-2を含む。

使用者辞書は機械翻訳サービスに新たな使用者が加入するとき作成されることが始動される。相互作用モードでは、新たな使用者には、翻訳のために提出されると期待される代価的なサンプルばかりでなく使用者に関連する用語の辞典のための個別化した単語を提供することが勧められる。その後、使用者保守モジュール13b-1の入力編集器は使用者の辞書の保守に貢献する基本的な情報を与えるように使用者に勧める。入力編集器は色々な文脈に関連してもよい。例えば、翻訳セッション中、入力編集器は、入力テキ

スト内の知らないスクラッチ単語または句に出会った際、使用者辞書に対するスクラッチエントリーを作成してもよい。辞書作成セッション（翻訳セッションの外側で）中、入力編集器は使用者によって供給されたリストからスクラッチエントリーを作成してもよい。辞書維持セッション中、入力編集器は確認と更新のために個人的な辞書の文脈を使用者に与えてもよい。3つの文脈のすべてにおいて、入力編集器は、使用者に言語上の原理や辞書データベースの構造の最少の知識しか必要としないで、使用者に関連する1つまたはそれ以上のドメインに入れられた単語の関係に対する使用者の知識を抽出するように試みる。

使用者には、相互作用が精巧である程度を選択することが与えられる。例えば、使用者は新たな単語の統語的および語彙的特性について入力編集器が与えた質問に回答することを選択してもよい。他の方法としては、使用者は単語の基本的な翻訳に不可欠な言語上の情報だけを与えるように設計された略字のオプションを選択してもよい。使用者には、また、特定の単語に対する新たなエントリーを作成しないが、許容できる代わりの単語または表現を提供することまたはソース単語を翻訳されないターゲットテキストに転送することを決定するオプションが与えられてもよい。これらの選択のすべてがDMOによって後に参考にするためにセーブされる。

第5図に示すように、スクラッチエントリーの作成を要求する相互作用は辞書データベースにある同様なエントリーまたは同義語を参照してもよい。知らない単語または句に出会ったとき、入力編集器は、単語がドメイン特定用法であるか否かを使用者に尋ねてもよい。もしそうなら、使用者には適当なドメインに名前を付けるのを勧められ、または置べき辞書データベース内に確立

されたドメインのリストが与えられる。次に、入力編集器は知らない単語の同義語を使用者に勧める。適当なドメイン内のサブ言語辞書がその同義語に対するエントリーのために検索される。もし1つの同義語が関連したサブ言語辞書に存在するならば、それが使用者辞書に入れられる。もしなければ、使用者には他の同義語が勧められ、このプロセスが繰り返される。もし同義語がドメイン特定辞書内に見出されなければ、コア辞書が同義語に対して検索されてもよい。

もし適当な同義語が見出されたならば、新たなエントリーの割当て名前としての新しい単語と、文脈としての同義語エントリーの文脈とを用いて、「コピーキャット」エントリーが使用者辞書内の新しい単語に対して作成される。使用者には、文脈「ass is」を用いる選択が与えられてもよく、単語が「コピーキャット」エントリーの文脈によって特定された方法で翻訳される。もし使用者がエントリー「ass is」を用いたくないなら、入力編集器は、エントリー文脈がコピーされた同義語に対して統語的および語彙的特性に関して同様であるか否かを強調するために、新しい単語の統語的および語彙的特性に関する情報を使用者に勧める。相互作用が新しい単語の仕様が可能にするならば、単語は翻訳されてもよい。もしそうでなければ、スクラッチエントリーがDMOによって後に参照されるために維持され、使用者は代用単語または表現を与えることを選ぶことができ、ソース単語を翻訳されないターゲットテキストに入れる。

第7A図～第7D図は、使用者辞書内のスクラッチ単語エントリーの作成とより高級なサブドメインまたはドメイン辞書への昇格の例を示す。第7A図は、使用者Bの辞書内の同様なエントリーに基づいて、使用者Aの辞書内に単語「マウス」（コンピュータの

周辺機器）に対するエントリーの作成を示す図である。両方の使用者はサブ言語「コンピュータ」の一員である。両方の使用者の辞書は言語対英語-日本語のコア辞書内に入れ子にされている。第7A図は転写形式の機械翻訳装置に対するプロセスを示す。プロセスは均等な方法で中間言語装置に対して実行されてもよい。

使用者Bの辞書は、転写に必要な3つの形式のエントリー、即ち英語モノリンガルエントリー（「E-単語」と表示）、英語-日本語バイリンガル転写エントリー（「EJ-単語」と表示）、および日本語モノリンガルエントリー（「J-単語」；「コンピュータマウス」と表示）を含む。これらのエントリーは単語の文法上の機能と言語上の特徴に属する規則および特徴に基くするために特に表示した仕様を含む。文法上の機能は言語上の形式（名詞、動詞等）のための一般的な規則に基くするフレーム参照符によって特定される。そのような参照符を含むことによって概してMT装置による使用に利用されるファイルからの承継による関連する翻訳処理規則が求められる。言語上の特徴は個々の単語に特有な特徴であり、それらの値はエントリー自身内に供給される。

例えば、第7A図で示す英語モノリンガルエントリーが名詞形成のフレーム参照符（E-N形式-1）および3つの特徴（E-N特徴-X等）を含む。バイリンガルエントリーは転写規則フレーム参照符（EJ-N形式-M）および単語「マウス」（日本語の翻訳はここでは**で表示される）の翻訳の仕様を含む。日本語モノリンガルエントリーはフレーム参照符（J-N形式-R）および3つの特徴を含む。2つのモノリンガルエントリーはグループ仕様、即ち、グループ=c o m（コンピュータ）を含む。

使用者Aによって装置のテキストに入力される新しい単語「マウス」に出会うと、入力編集器は次のように使用者に相互作用す

る。

質問：これはドメイン特定方法ですか？

回答：はい

質問：ドメインはなんですか？

回答：コンピュータです。

この情報で、入力編集器は「コンピュータ」サブドメイン内の他の使用者の辞書进行操作し、使用者Bの辞書内の「マウス」に対するエントリを見出す。入力編集器は、グループ形式、例えば、「マウス」に対する使用者Bのエントリが使用グループ＝comを含むか否かを含めて、このエントリが実際に質問のドメインに属するか否かを確認するために、チェックを行う。単語「マウス」に対するグループ仕様が、使用者Bが他のグループおよびドメインの一員であるかもしれないので、サブドメイン「コンピュータ」内に使用者Bが属することに加えて、必要であるかもしれない。使用者Aに対してなされる他の質問の書式でエントリの関連を決定するために付加的なチェックがなされてもよい。この図示の例では、例えば、入力編集器は使用者Aに、単語「マウス」が名詞（フレーム参照行名前およびエントリ内の特徴から名詞状態を推論させる）であるか否か、また単語がN特徴X等に対して正値を付するかを質問してもよい。

使用者Bの辞書に見出された「マウス」に対するエントリが使用者Aの目的に適切に合ったものであることを決定する際、使用者Aの辞書に対するエントリが使用者Bの辞書からの関連するエントリのコピーを作ることによって作成され、トランザクションの記録がDMOによって後に参照されるために作られる。コピー処理によって、名詞の一般的な機能ばかりでなく単語「マウス」に特有な特徴がフレーム参照符（第7A図の上部中央内の小さい

ボックスによって図示されている）を通した使用者Aのエントリによって引き継がれる。

もし単語に対するエントリが「コンピュータ」ドメイン内の使用者の辞書に見出されないならば、入力編集器はコア辞書も含めて、より低級またはより高級レベルの辞書内のエントリに対して走査してもよい。検索が使用者Aによって特定されたドメインからさらに移動すると、入力編集器は、使用者に対するさらの別のチェックと通知によって使用者のスクラッチエントリを作成するために候補のエントリの選択をする際により大きな注意を促し、DMO記録内の警告の表示を含むことができる。第7C図は、同一のエントリが他のドメイン内の他の使用者の辞書に見出された例を示す。

もし同一の単語に対するエントリが利用している辞書に見出されなかったならば、入力編集器は単語の同義語を要求してもよい。この例の場合、使用者Aは「マウス」に対する同義語として「指示器」に回答してもよい。次に、装置は同義語のエントリに対する種々のレベルの辞書を走査できる。もし同義語エントリが、例えば、使用者Fの辞書に見出されたならば、入力編集器は前述したように見出したエントリの特徴に関して使用者Aに質問することができる。もし使用者Fの同義語のエントリが使用者Aの目的に対して適切に適合することが決定されると、使用者Aの用語（例えば、「マウス」）のエントリレベルと「指示器」に関する使用者Fのエントリを持った「コピーキャット」エントリが作成される。再び、このトランザクションの記録がDMOによるさらに別の監視のために作られる。第7D図は「コピーキャット」エントリプロセスの図を与える。

使用者辞書に使用者保守モジュール13b-1の入力編集器を

通してスクラッチエントリを付加することによって、言語上の特徴およびドメイン（サブ言語）の関係に関する基本情報が供給される。予め仕切られた文章の集成がMT装置に与えられる。予め仕切られた文章の集成は新たな単語と用法の高品質のソースを与え、その結果、言語学者は言語用法の分野の調査に対する努力が節約される。辞書データベースはテキストで出会うおよび（または）使用者によって供給される実際の単語および用法に沿って絶えず更新される。

本発明の重要な特徴は、単語エントリがより高級な辞書に含まれることが望ましいと指示するような、言語上の完全性とより一般的な用法であることを確認する或る種の試験を満足させるときには、より低級なレベル（使用者、サブ言語/グループ、ドメインまたはコア）の辞書からより高級なレベル（サブドメイン/グループ、ドメインまたはコア）の辞書に単語エントリを移動する能力である。言語は、新たな単語と用法が個人およびグループによって採用され、次いで、より大きなグループおよび社会全体に流行するので、絶え間ない発展の状態にある。より大きな辞書へのエントリの移動は、不正確な翻訳または辞書データベースの崩壊が生じないようにするために、言語的に関係された辞書保守動作器（DMO）および新たな単語と表現の翻訳に基づく検討と監視が必要である。本発明は、より高級なレベルの辞書への単語エントリの移動の際にDMOを支援するDMO保守モジュール13b-2内に自動化したユーティリティを与える。もし言語的および統語的特徴がプログラム制御の下に信頼できるように決定されるならば、DMO支援ユーティリティは自動化した基盤に基づきそのような単語のエントリの移動を行うことができる。DMO支援ユーティリティは、機械翻訳装置の正確さ、性能、有用性を改

良できる非常に大きな辞書データベースのほぼコンピュータ化した処理の方法を提供する。

辞書保守のマニホールドタスクがこの明細書に記載された機械翻訳装置の動作の基本である。DMOのタスク（仕事）は新たな辞書エントリを加えことおよび既にあるエントリを削除または変更することを監視することである。エントリが加えられるか削除されるか、エントリのどの特徴が変更されるか、辞書およびサブ言語辞書がこれらの変更によって影響をうけるかに関する選択は、プログラムされた支援ユーティリティを採用することによって誘導される用法上のデータとエントリの効率に基づいてDMOによってなされる。

例示として、DMO保守モジュール13b-2は下記の集成からの単語エントリを抽出して構成するのを支援する基本ユーティリティを含んでもよい。即ち、精査と手動編集を容易にするために、編集された辞書エントリ（文法規則も含む）を表示用のテキストファイルに変換すること、前述の記録の自動化した作成およびテキストファイルとして記録を表示すること、辞書エントリの構造上の誤りの検出およびテキストファイルとして誤ったエントリを表示すること、エントリ名だけによる辞書エントリのすべてのリスト（辞書内のエントリの「シェルフリスト」）を抽出して表示すること、エントリ名および翻訳だけ（明瞭さに対する統語上のカテゴリーを含む）によるバイリンガル辞書エントリのすべてのリスト、即ち、バイリンガル「シェルフリスト」を抽出して表示すること、および（または）そのようなリンクなしにより高級な辞書にアクセスすることを防止するために引き継ぎ階層内の誤ったリンクを検出することを含むものである。

抽出と表示ユーティリティのセットは辞書データベースの監視

において、一貫性と制御のレベルを維持する際DMOを支援できる。本発明の重要な焦点は、使用者のグループの辞彙および言語的な特有の表現に適したサブ言語を識別し、作成し、維持し、用いることである。これらの態様は、同一のドメインまたはグループの一員の流動的な用語の用法パターンに基づいて、サブ言語間の関係を進行する評価と保守を必要とする。

これらの辞書関連のタスクを実行する際に最大の効率を得るには、DMOに利用できる精巧なユーティリティが要求される。そのようなユーティリティは同一のドメイン内の使用者による新たな単語のエントリのすべての場合に追従する能力を含んでもよい。使用回数ユーティリティは使用の回数、エントリの数、好ましい同義語のエントリの識別を決定できる。他のユーティリティは使用者辞書からドメイン辞書への可能な昇格のためにDMOにそのようなデータを与えることができる。サブ言語選択ユーティリティは、サブ言語辞書選択プロセスを正確にするために、前述の記録の分析を行い、個々の使用者およびグループの使用者によるパターンを表示することができる。同形異義語エントリユーティリティは、同形異義語のエントリの最適な形式を決定する際に分析を行い、より高級なレベルの辞書にエントリを昇格させるようにDMOに指示するために、所定のドメイン内の使用者辞書内の同形異義語のエントリを識別し表示する。品質補償ユーティリティは品質補償チェックのためにまたは許容したエントリのより高級なレベルの辞書への昇格のために新たなスクラッチエントリを生成できる。

装置の言語上および語彙上の特徴構成の一貫性を維持するのを支援するものとして、ユーティリティは、また、或る特徴値対を含むエントリ、例えば特徴「+液体」を含むエントリを見出し

表示し、或る特徴値対を含むエントリのシェルリストを表示し、特徴評価の順りに追従するのを支援するように特徴の構成を表示するために、設けられてもよい。

DMO支援ユーティリティは、異なった辞書間の整合性、矛盾または二重エントリを解決するために、引用辞書、比較、一致、集積、逆、統計的整合性、集団分析を行うためのアルゴリズムを用いてもよい。例えば、ユーティリティは、任意の単語エントリがより一般的な単語エントリに入れられて使用者辞書からドメイン辞書に昇格されてもよいか否かを見るために、同一のドメインまたはグループから使用者の辞書を走査するのに用いられてもよい。そのようなエントリ昇格の例が第6図に示されている。

低級な辞書内のエントリ間の類似性の充分なレベルの判断基準は、前述のような統計的なまたは数値のアルゴリズムを用いて、検査される。このような検査値や他の値は、エントリのどの特徴が一般的であり、したがって、より高級なレベルの辞書のために作成されるエントリに含まれることが適したものであるか、また使用者の辞書がエントリのソースである使用者に特有であるかを決定するのに用いられてもよい。より高級なレベルのエントリは適用できる一般的な特徴をもったものだけに関して行われる。特有の特徴は、もしあれば、使用者辞書に保持されてもよい。エントリの昇格は辞書の階層の隔ったレベル間で生じる。

DMO支援ユーティリティは、また、グループ構成員およびグループまたは構成員間の関係を決定し、同様なグループまたは構成員に対して記録された単語および用語に基づいて特徴を推論し、またはマスター（スーパーオーディネート）サブ言語辞書の作成のためのサブ言語用語を誘導するために、個々の使用者または使用者のグループの入力テキスト内の単語および用語の同時発生を

検査してもよい。そのような分析は、或る使用者に適切なサブ言語辞書を選択するのに用いられる辞典用のキーワードのリストを提供する。

DMO支援ユーティリティによって支援された情報によって、DMOは1つのサブ言語辞書に入れ、1つのサブ言語辞書から削除し、1つのサブ言語辞書から他のサブ言語辞書に移動させるための単語および句のリストを作成できる。同様に、DMOはエントリの適切なリスト内の特徴を変更するために特徴抽出ユーティリティによって供給されるデータを用いてもよい。DMOは、データをテキストファイルに入力し、ファイルを辞書データベース内に構築することによってエントリを作成し変更してもよい。しかしながら、辞書内のソースエントリ、ターゲットエントリおよびソースからターゲットへのエントリを含む辞彙エントリの作成および変更を自動化するユーティリティの一層精巧で効率的なセットが設けられてもよい。

要約すると、辞書制御モジュールによって機械翻訳装置全体は非常に流動的で高い柔軟性のサブ言語能力を持つことができる。サブ言語能力は、実際の使用者、使用者グループ、ドメインまたは分野の出来た単語および識別された選択に基づいて、開発された時間がたつにつれて蓄積される。多数のサブ言語辞書は、エントリが用法に沿って変更でき、ドメイン特定用法に対する容量が実際上制限がないことを除いて、実際の辞書内の同義語および他の句の用法のリストに似たものである。多数のサブ言語辞書の操作におけるコンピュータの力は、所定のドメインに対するサブ言語パターンのセット全体を厳密に定義する要求に置き代わる。したがって、サブ言語能力の水平方向の拡張が垂直方向の定義に置き代わる。

出力の書式化および送信

第1C図に示すように、出力モジュールは、認識モジュール12によって認識されたテキストのためにカバーページまたは見出しから抽出された指示を書式化することによって駆動されるページ書式化（フォーマット化）能力部31と、カバーページから抽出され認識モジュール12によって供給される住所にいる受取人に遠隔通信リンクを介して書式化した出力テキストを送信する送信インターフェース32を含む。

一旦入力テキストがターゲット言語テキストに翻訳されると、出力モジュール30のページ書式化能力部は、カバーページに指定された書式化情報に基づいて所望のページ書式に翻訳されたテキストを編集する。例えば、英語-日本語翻訳に対しては、出力日本語テキストはカナの左から右への水平ラインまたは表意文字の右から左への垂直ラインとして書式化される。ページ書式（フォーマット）は、また、ページ毎の翻訳が指定されるかもしれない。その場合、書式化プログラムはソースおよびターゲットテキストの圧縮比を考慮する。例えば、第3図に示すように、英語のテキストは表意テキストよりも一般にはより広がっており、英語のテキストの8.5インチ×11インチの入力ページは、通常により大きなポイントの寸法の中国文字とライン間スペースを持って同一寸法のページに再書式化されてもよい。これに対応して、表意文字テキストの代表的な15.2cm×25.6cmのページが英語テキストの8.5インチ×11インチのページに再書式化されてもよく、A4サイズのページが8.5インチ×11インチページとして再書式化されてもよい。

書式化プログラムは、また、第3図の部分Fに示すように、脚注機能を実行してもよい。脚注機能は、原稿のソース言語テキス

トを復写することによって入力テキストの不明瞭な単語または句に対する脚注（例えば、単一のアスタリスクで示す）を与え、および（または）ターゲット言語の代わりの翻訳（二重アスタリスクで示す）を与える。ソース言語テキストおよび（または）代わりの翻訳は、翻訳処理中解決できない不明瞭な単語または句に対してフラッグを立てることによって与えられる。他の周知の書式化機能、例えば、余白、ページレイアウト、脚、翻訳不可能な図形像の復写等は出力モジュール30によって行われてもよい。

書式化した出力文書が出力送信状態にあるとき、送信インターフェース82は対応する出力装置を制御して遠隔通信リンクBを介してそれぞれの装置に電子データ信号として出力文書を送信する命令信号を発生する。出力装置は、電話、ファクシミリ/モデムボード、プリンタ、ファクシミリ装置または出力データをネットワーク上の受取人電子住所に送信するネットワークインターフェースを含んでもよい。第2図に示すように、カバーページは異なった言語の異なった住所に位置する複数の受取人を指定してもよい。送信インターフェース82は出力データの適切なフォームを発生し各受取人に送信する。例えば、もし各受取人がファクシミリ送信とプリントしたコピーを受け取ることを指定したならば、送信インターフェースはデータをファクシミリ/モデムボードを通して送信しかつ郵便用ハードコピーをプリントするプリンタを動作させる。

一般的な遠隔通信の適用性

前述の機械翻訳装置は、例えば、私的なおよび公的なネットワークおよびゲートウェイ会社、遠隔通信会社およびバイリンガルまたはマルチリンガル情報サービス提供者のような駐在ユーティリティまたはサービスとして設置できる。装置への入力は好まし

くは類似用語に対する電子テキストの形態である。さらに他の開発によって、グラフ（ファクシミリ）データおよびスピーチの形態の入力は捕捉され、走査され、および（または）翻訳処理のための中間テキストに変換され、次に任意の所望のフォームの出力に変換される。辞データベースは広範囲なドメインに対してサブ言語の範囲の大きさ内で開発されるとき、装置は広範囲の分野にわたって通信を取り扱う正確さと能力を兼ね備える。多量の配分量、安価な処理力と速度は、多数の言語の翻訳に対して、多くの異なった言語対、技術分野、ドメイン、使用者グループ、使用者を取り扱うのに有効に利用される。

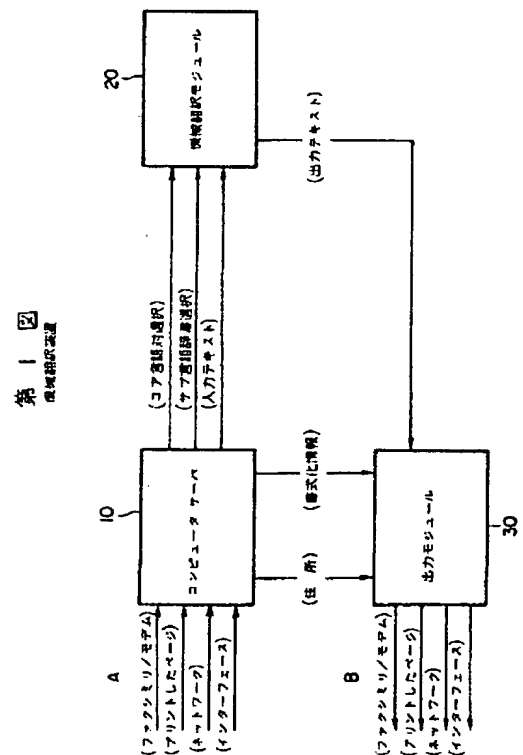
類似の用語の使用に対して、機械翻訳装置は、電子メール、電子ブルテンボード、情報、ネットワークサービスに対して、電子テキストを翻訳するのに特に適している。前述のように、相互作用モードがネットワークを介して翻訳され送信されるオンラインの使用の入力テキストと相互作用するために設けられてもよい。このモードで、使用者には、カバーページの領域を書き込み、使用者IDファイルを作成し保守し、または辞書データベースを更新することが勧められる。

異なった言語における電子テキスト入力および出力に対して、世界中にわたって用いられる多数の異なった文字コード規約に対する標準化したインターフェースを持つことが望ましい。世界的な文字コード規約がカルフォルニア州、マウンテンビュー所在のユニコードコンソーチウムによって結ばれた。コンソーチウムにはIBM、DEC、アップルおよび他の主要なアメリカのコンピュータ会社が入っている。ユニコードのセットは、主要なローマアルファベットシステムおよびアジアの文字システムを含む世界の主要な文字コード規約にマッピングされた16ビット文字コー

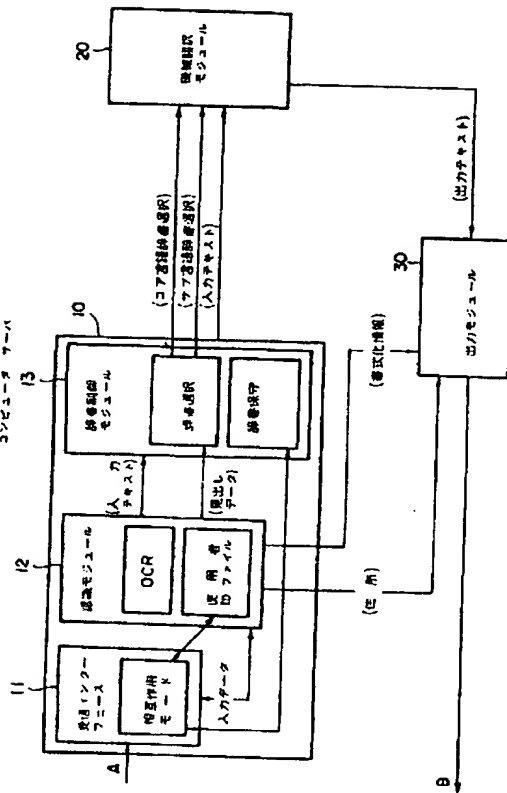
ドセットである。例えば、ユニコードセットは中国、日本、韓国、台湾で用いられている主要工業および国民の標準のハン文字セットにマッピングされている。このように、ユニコード文字変換モジュールは遠隔通信装置内の電子テキスト用の標準化したインターフェースとして用いられることができる。

テキスト入力は、プリントしたページからまたはファクシミリ/モデムを介する送信から走査される。装置の認識装置はそのような走査したページの像データを機械読取り可能なテキストに変換するのに用いられる。最近では、オフザシェルフプログラムが英語の英数字と日本のカナに対して利用できる。他の文字セット、例えば中国語のハンおよび日本語の漢字文字、さらに手書きの文字に対する文字認識プログラムの将来の開発がページ指向の翻訳装置の開発に期待されている。

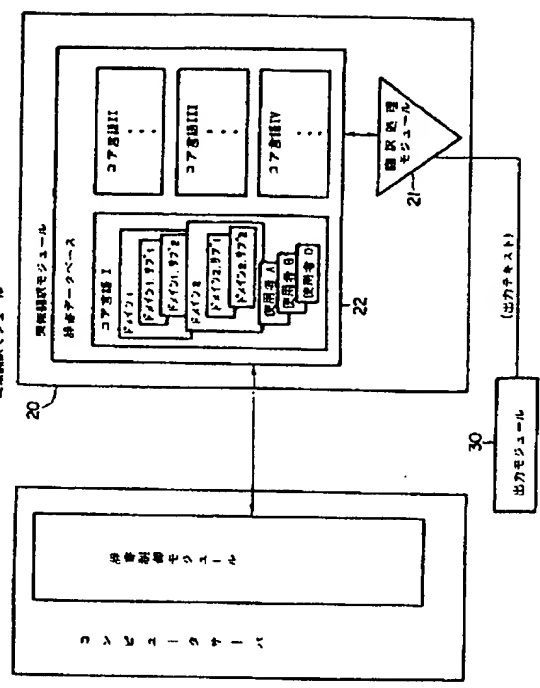
第1図
機械翻訳装置



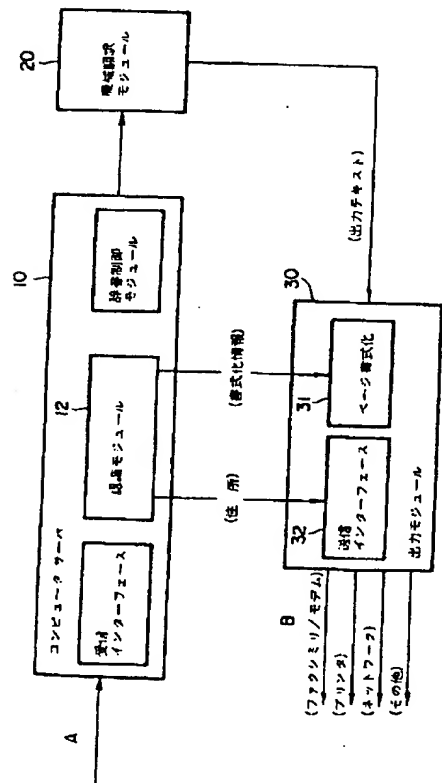
第 1A 図
コンピュータアーバ



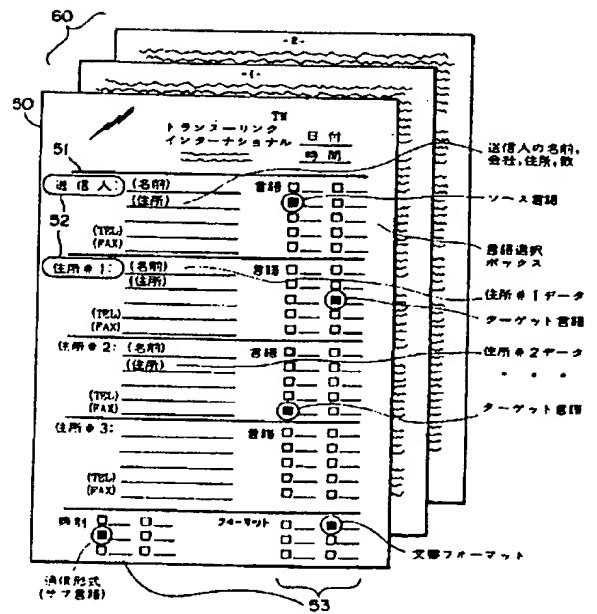
第 1B 図
通信制御モジュール

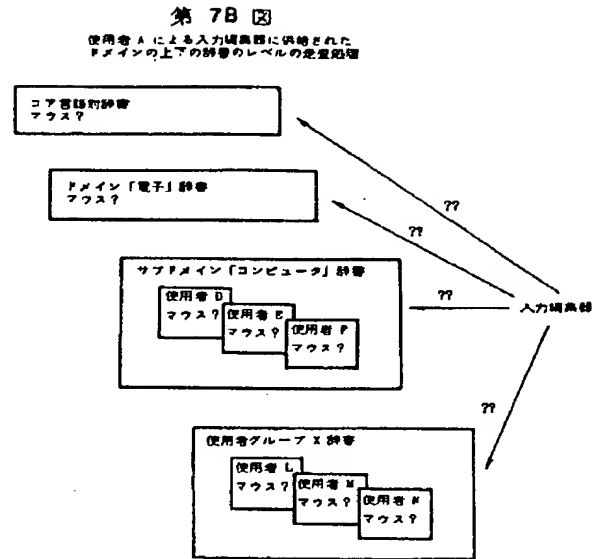
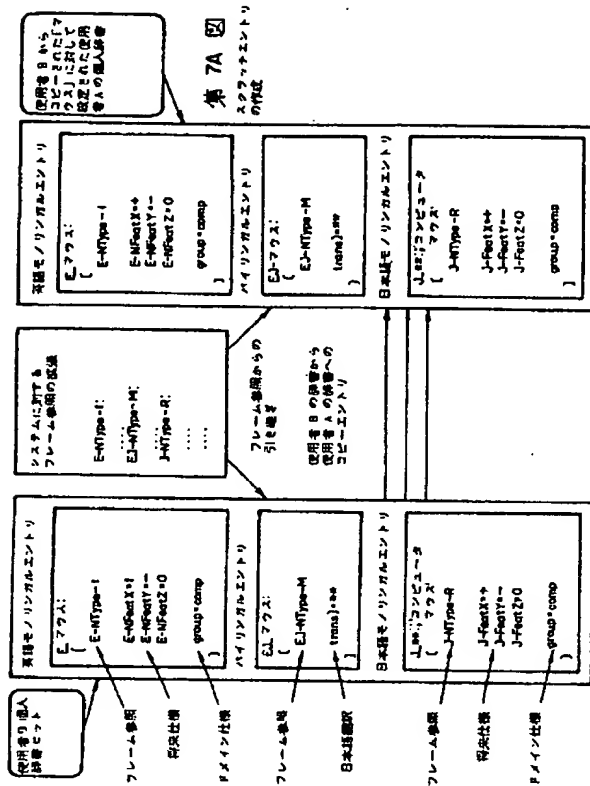


第 1C 図
出力モジュール

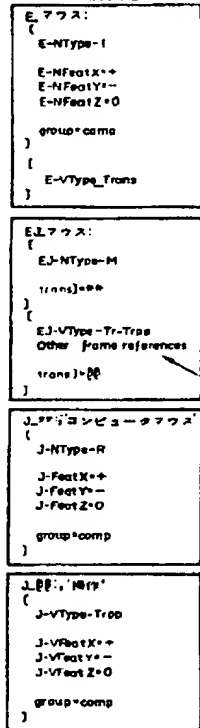


第 2 図

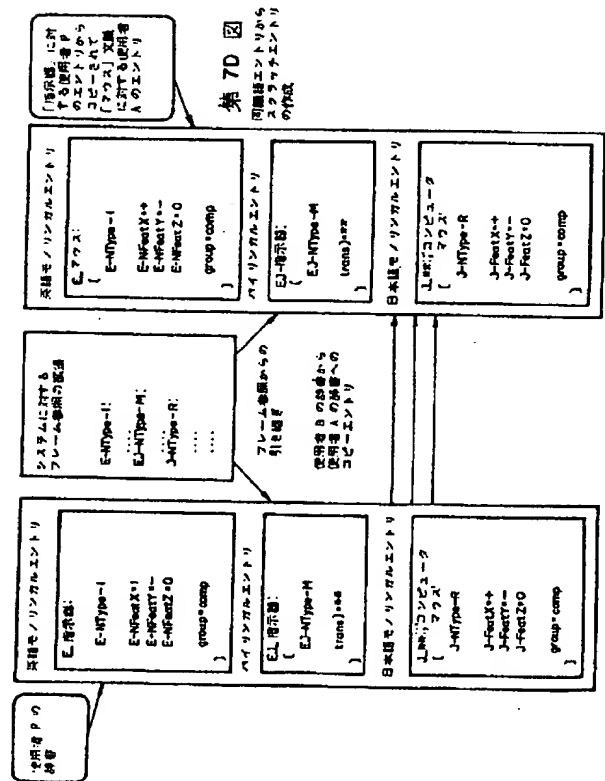




第7C図
「マウス」に関連するスクリーンエントリの作成



他のエントリおよび「マウス」=「マウス」を操作する」に関連する文法規則を参照する



手続完結正付

平成6年5月30日



国際調査報告

特許庁長官 殿

1. 国際出願の表示

PCT/US91/09761

2. 発明の名称

機械翻訳装置の通信装置

3. 補正をする者

事件との関係 特許出願人

氏名(名称) トランスリンク インターナショナル コーポレーション

住所 アメリカ合衆国 ニューヨーク州 10017

ニューヨーク 800 イースト 42nd ストリート
18ス フロア

国 籍 アメリカ合衆国

3. 代理人

氏 名 (8758) 弁護士 田中 埤郎

住 所 東京都千代田区神田麹町3-3-9

麹町ビル(新千代田) 73号

4. 補正命令の日付

平成6年2月28日(発送日平成6年3月8日)

5. 補正の対象

特許法第184条の5第1項の規定による書面、委任状原文及び捺印

6. 補正の内容

別紙のとおり

International Application No. PCT/US91/09761

IPC (3): 659 T8/38
U.S. Cl.: 364/419

U.S. Cl.: 364/419, 379/90

Documents cited by the applicant:

Category	Document	Date of Document	Relevance
E-X	US, A, 5,077,804 (RICHARD) 31 December 1991	31-12-91	1-4, 6, 8, 10
E-X	See figure 3, column 2, lines 29-40, column 6, lines 31-50, column 8, lines 39-47, column 12, lines 40-52, and column 13, lines 10-16.		3, 7, 9, 11-20
Y	US, A, 4,980,829 (OKAJIMA ET AL.) 25 December 1990	25-12-90	11-20
Y	See column 9, lines 20-24 and column 10, lines 39-50.		
Y	US, A, 4,352,012 (VERDERER ET AL.) 28 September 1983	28-09-83	1-10
Y	See figure 1.		
Y	Patent Abstracts of Japan, Vol. 14, No. 489, 10 August 1990, Abstract No. 02-202143, Hashimoto.	10-08-90	1-10
Y	See the entire document.		
Y	Patent Abstracts of Japan, Vol. 11, No. 319, 18 May 1987, Abstract No. 62-107376, Dagami.	18-05-87	1-10
Y	See the entire document.		
A	US, A, 4,383,307 (GIBSON, III) 10 May 1983	10-05-83	11-20
A	See figure 1.		

16 February 1992

08 APR 1992

David Hartley

International Application No. PCT/US91/09761

Documents cited by the applicant:

Category	Document	Date of Document	Relevance
A	Patent Abstracts of Japan, Vol. 12, No. 62, 11 September 1987, Abstract No. 62-203273, (HUKU ET AL.) See the entire document.	11-09-87	11-20
A	Patent Abstracts of Japan, Vol. 12, No. 107, 12 October 1987, Abstract No. 62-232087. See the entire document.	12-10-87	11-20
A	Patent Abstracts of Japan, Vol. 12, No. 187, 18 December 1987, Abstract No. 62-291250, (Kenbo et al.) See the entire document.	18-12-87	1-10
A	Patent Abstract of Japan, Vol. 12, No. 238, 08 February 1988, Abstract No. 63-29861, (Kobe et al.) See the entire document.	08-02-88	11-20
A	Patent Abstracts of Japan, Vol. 14, No. 249, 05 March 1990, Abstract No. 02-64767, (YODA) See the entire document.	05-03-90	11-20

International Application No. PCT/US91/09761

Documents cited by the applicant:

Category	Document	Date of Document	Relevance
A	US, A, 4,882,681 (BROTZ) 21 November 1989	21-11-89	1-10
A	See Figure 1.		
A	US, A, 4,916,730 (HASHIMOTO) 10 April 1990	10-04-90	1-10
A	See the entire document.		
A	Analyzing Language in Restricted Domains, 1986, Associated Pub., Hillsdale, NJ, Erlbaum et al., preface pages x-xviii.		11-20
A	Patent Abstracts of Japan, Vol. 11, No. 284, 16 April 1987, Abstract No. 62-828853 (NAJAKETAMA) See the entire document.	16-04-87	1-10

Observations where certain claims were found unsearchable:

The International Searching Authority has not searched for prior art in the following countries:

☐ China: because they are not part of the International Searching Authority's search area.

☐ Cuba: because they are not part of the International Searching Authority's search area.

☐ Japan: because they are not part of the International Searching Authority's search area.

☐ Korea: because they are not part of the International Searching Authority's search area.

☐ Mexico: because they are not part of the International Searching Authority's search area.

☐ Russia: because they are not part of the International Searching Authority's search area.

☐ Singapore: because they are not part of the International Searching Authority's search area.

☐ Taiwan: because they are not part of the International Searching Authority's search area.

☐ Thailand: because they are not part of the International Searching Authority's search area.

☐ United States: because they are not part of the International Searching Authority's search area.

☐ Vietnam: because they are not part of the International Searching Authority's search area.

【公報種別】特許法第17条第1項及び特許法第17条の2の規定による補正の掲載
 【部門区分】第6部門第3区分
 【発行日】平成11年(1999)9月14日

【公表番号】特表平6-508941
 【公表日】平成6年(1994)10月6日
 【年通号数】
 【出願番号】特願平4-503420
 【国際特許分類第6版】

G06F 17/28
 H04L 29/10

【F I】

G06F 15/38 Z
 C
 H04L 13/00 309 Z

手続補正書

平成10年12月24日

特許庁長官 殿

7. 補正の内容

請求の範囲を別紙の通り補正する。

1. 事件の表示

平成4年特許第503420号

2. 発明の名称

機械的変換装置の改良

3. 特許出願人

姓 名 アメリカ合衆国 ニューヨーク州 10017
 ニューヨーク 300 イースト 42nd
 ストリート 18ス フロア

氏 名 トランスレーンク インターナショナル コーポ
 レーション

通 称 アメリカ合衆国

4. 代理人 平101-0044

住 所 東京都千代田区錦糸町1丁目6番15号
 丸ビル(春日駅側)32号
 電話(03)3258-0183

氏 名 (3758)井原 田中 雄三

5. 補正により増加する請求項の数 9

6. 補正の対象

請求の範囲

図 2 の 説 明

1. 機械翻訳装置の構成図において、
 - (a) ソース言語の入力テキストからターゲット言語の出力テキストに機械翻訳を行う機械翻訳モジュールと、
 - (b) 翻訳された出力テキストが送達される受取人の住所を指定するのびくとも第1所定の領域を持つ制御部分を持った翻訳すべき入力テキストを含む入力を受け、送達送信リンクを介して送信する受信インターフェースと、
 - (c) 制御部分を単元的に制御し制御部分の第1所定領域に指定された受取人の住所を認識するために受信インターフェースに接続された認識モジュールと、
 - (d) 機械翻訳モジュールによって翻訳された出力テキストを認識モジュールによって認識された受取人の住所に照らして送達送信リンクを介して送達するための送信インターフェースを含む出力モジュールと、
2. 請求の範囲第1項記載の機械翻訳装置において、受信インターフェースは光信号、音声信号、ファクシミリ信号およびネットワークから送る群のうちの1つまたはそれ以上の信号を通して入力を受信するように第1送達送信リンクに接続されていることを特徴とする機械翻訳装置送達送信装置。
3. 請求の範囲第1項記載の機械翻訳装置送達送信装置において、受信インターフェースへの入力はテキスト文字を有するデータであり、認識モジュールがデータ機械翻訳可能なテキスト文字に接続する文字認識手段を有することを特徴とする機械翻訳装置送達送信装置。
4. 請求の範囲第1項記載の機械翻訳装置送達送信装置において、機械翻訳モジュールは装置のソース／ターゲット言語にそれぞれ

- 対応する辞書の二つの辞書を含み、制御部分は翻訳すべきテキストに適用可能な選択したコア辞書を指定する第2所定部分を有し、認識モジュールは制御部分の第2所定領域内の指定を認識して機械翻訳モジュールに選択したコア辞書辞書を使用させる辞書制御手段を有することを特徴とする機械翻訳装置送達送信装置。
5. 請求の範囲第1項記載の機械翻訳装置送達送信装置において、機械翻訳モジュールは辞書のサブ辞書辞書を有するデータベースを有し、制御部分は翻訳すべきテキストに適用可能な第3の所定部分を有し、認識モジュールは制御部分の第2所定領域内のサブ辞書辞書を認識し機械翻訳モジュールに選択したサブ辞書辞書を使用させる辞書制御手段を有することを特徴とする機械翻訳装置送達送信装置。
6. 請求の範囲第1項記載の機械翻訳装置送達送信装置において、受信インターフェースはフックンリプログラム、ネットワークインターフェース、およびプリンタから送る群のうちの1つまたはそれ以上の信号を通して出力を送信するために第2送達送信リンクに接続されていることを特徴とする機械翻訳装置送達送信装置。
7. 請求の範囲第1項記載の機械翻訳装置送達送信装置において、出力モジュールは制御部分の他の所定領域内で指定された選択したページ形式に機械翻訳モジュールからの翻訳された出力テキストを形式化するページ形式化モジュールを有することを特徴とする機械翻訳装置送達送信装置。
8. 請求の範囲第1項記載の機械翻訳装置送達送信装置において、受信インターフェースが第1送達送信リンクを介して受信インターフェースを有する使用人方向のプログラムされた相互作用モジュールを有することを特徴とする機械翻訳装置送達送信装置。
9. 請求の範囲第8項記載の機械翻訳装置送達送信装置において、相互作用モジュールは使用人によって使用される制御部分のため

- にデータ配線を接続する使用人識別ノードに接続されており、データ配線は使用人によって選択され、翻訳すべき入力テキスト用の所定部分のために使用されることを特徴とする機械翻訳装置送達送信装置。
10. 請求の範囲第1項記載の機械翻訳装置送達送信装置において、機械翻訳モジュールは受信インターフェース、認識モジュールおよび出力モジュールから分離して翻訳を行う制御プログラムされた処理手段を有することを特徴とする機械翻訳装置送達送信装置。
11. 機械翻訳装置において、
 - (a) 入力テキストと機械の可能なサブ辞書辞書から入力テキストに適用可能な選択したサブ辞書を指定するサブ辞書制御入力を受信する受信インターフェースと、
 - (b) ソース言語の単語に列挙するターゲット言語の単語に列挙するエントリを含む辞書データベースを用いてソース言語の入力テキストのターゲット言語の出力テキストへの機械翻訳を行う機械翻訳モジュールと、
 - (c) ソースおよびターゲット言語の二語単語に対するエントリを含むコア辞書辞書とサブ辞書の特殊な単語のエントリを含む辞書のサブ辞書辞書を有する辞書データベースと、
 - (d) 入力テキストに適用可能な辞書データベースのサブ辞書辞書を認識し機械翻訳モジュールに入力テキストの翻訳のための選択したサブ辞書辞書を指定するためにサブ辞書制御入力に接続する辞書制御モジュールと、
 - (e) 機械翻訳モジュールによって翻訳されたターゲット言語のテキストを出力する出力モジュールと、
12. 請求の範囲第11項記載の機械翻訳装置において、辞書データベースは機械のソース／ターゲット言語にそれぞれ対応する辞

- 書のコア辞書辞書を有し、受信インターフェースは入力テキストに適用可能な選択されたコア辞書辞書を指定するソース／ターゲット制御入力を受信し、ソース／ターゲット制御入力に照準する辞書制御モジュールが機械翻訳モジュールに入力テキストの機械のために選択したコア辞書辞書を使用させることを特徴とする機械翻訳装置。
13. 請求の範囲第11項記載の機械翻訳装置において、受信インターフェースは入力テキストに照準する使用人を識別するデータを有するサブ辞書辞書入力を受信し、辞書制御モジュールは受信インターフェースによって受信した使用人識別データに従ってテキストに適用可能なサブ辞書辞書を指定するための辞書選択モジュールを有することを特徴とする機械翻訳装置。
14. 請求の範囲第11項記載の機械翻訳装置において、辞書選択モジュールは送信人および受取人の資格、性別、会社、地位、色、使用言語のグループおよび辞書を含む使用人識別データから引当される推薦に基づいて適用可能なサブ辞書辞書を指定する推薦アルゴリズムを有することを特徴とする機械翻訳装置。
15. 機械翻訳装置において、
 - (a) ソース言語の単語に列挙するターゲット言語の単語に列挙するエントリを含む辞書データベースを用いてソース言語の入力テキストのターゲット言語の出力テキストへの機械翻訳を行う機械翻訳モジュールと、
 - (b) ソースおよびターゲット言語の二語単語に対するエントリを含むコア辞書辞書と使用言語のグループによって用いられるサブ辞書の特殊な単語のエントリを含む辞書のサブ辞書辞書を有する辞書データベースと、
 - (c) 機械翻訳モジュールにコア辞書辞書、任意の選択したサ

ブ管理符号、および使用者に既する入力テキストの照会のための任意の選択した使用符号群を使用させるために機械翻訳装置に列する制御入力に相当する符号制御モジュールと、

を有することを特徴とする機械翻訳装置。

15. 請求の範囲第15項記載の機械翻訳装置において、受信インターフェースはスタック単位エントリを生成する使用符入力のためのプログラムされた組立用モジュールを有し、符号制御モジュールは対応する使用符号にスタック単位エントリを加える使用符保守ユーティリティを有することを特徴とする機械翻訳装置。

17. 請求の範囲第16項記載の機械翻訳装置において、符号制御モジュールは他の使用符の群内の限られた基元エントリを参照して生成されるスタック単位エントリを生成する使用符保守ユーティリティを有することを特徴とする機械翻訳装置。

18. 請求の範囲第15項記載の機械翻訳装置において、符号データベースは入れ子形式の構造を有し、優先度レベルが最も低級なレベルにあり、より高次のレベルのリブ記号群内に入れ子にされており、サブ言語群は最も高次のコア言語群に入れ子にされており、符号制御モジュールは、機械翻訳モジュールによって、最も低級なレベルから最も高次のレベルの順に解釈された使用符群、生成されたサブ言語群およびコア言語にアクセスすることを特徴とする機械翻訳装置。

19. 請求の範囲15項記載の機械翻訳装置において、符号制御モジュールは符号保存装置(DMR)に単語エントリがドメインまたは使用者の使用符号に共通に与えられるかを検索させ、そのような単語エントリをそのドメインまたは使用符のグループ用のより内嵌なレベルのサブ言語群に移動させる符号保守ユーティリティを有することを特徴とする機械翻訳装置。

ゲット言語対の特定化した用語のサブ言語を指定する第1所定フィールドを含み、機械翻訳装置モジュールは各ソース/ターゲット言語対の対照辞書とソース/ターゲット言語対内のそれぞれサブ言語ドメインに適用可能な特定化した用語を有する複数のサブ対照辞書を有し、機械翻訳装置モジュールは対応するサブ言語群を識別し機械翻訳装置モジュールが出力テキストの翻訳を実行する際に選択したサブ言語群を用いるために制御入力中のリブ記号指定に相当することを特徴とする機械翻訳装置。

23. 請求の範囲21項記載の機械翻訳装置において、制御入力テキストおよび翻訳した出力テキストは電子テキストとして通信リンクを介して送信され、翻訳出力モジュールは翻訳された出力テキストを通信リンクを介して送信することを特徴とする機械翻訳装置。

24. 請求の範囲21項記載の機械翻訳装置において、制御入力テキストは通信リンクを介して送信され、翻訳出力モジュールは翻訳された出力テキストを通信リンクを介して送信することを特徴とする機械翻訳装置。

25. 請求の範囲21項記載の機械翻訳装置において、制御入力テキストおよび翻訳した出力テキストは電子テキストとして通信リンクを介して送信され、翻訳出力モジュールは翻訳された出力テキストを通信リンクを介して送信することを特徴とする機械翻訳装置。

26. 請求の範囲21項記載の機械翻訳装置において、制御入力テキストは通信リンクを介して送信され、翻訳出力モジュールは翻訳された出力テキストを通信リンクを介して送信することを特徴とする機械翻訳装置。

27. ソース言語からターゲット言語に自動的に翻訳して、翻訳

28. 請求の範囲第15項記載の機械翻訳装置において、符号制御モジュールは単語エントリがドメインまたは使用者の使用符号群に共通に与えられるかを自動的に検索し、そのようは単語エントリをそのドメインまたは使用者のグループ用のより高次のレベルのサブ言語群に移動させる符号保守ユーティリティを有することを特徴とする機械翻訳装置。

29. 請求の範囲第15項記載の機械翻訳装置において、

(a) ターゲット言語に翻訳された出力テキストを受信する受取人のアドレスを指定する第1所定フィールドと入力テキストの機械翻訳装置のソース/ターゲット言語対の中から選択したソース/ターゲット言語対を指定する第2所定フィールドを含む制御入力を作ったソース言語の入力テキストを通信リンクを介して送信する受信インターフェースと、

(b) 対応するソース/ターゲット言語対の選択の際、各々がソース言語の入力テキストからターゲット言語の出力テキストへの機械翻訳装置を有する機械翻訳装置モジュールと、

(c) 機械翻訳装置モジュールが対応するソース/ターゲット言語対のサブ言語群を用いて入力テキストの翻訳を実行させるために制御入力中のソース/ターゲット言語対の指定の符号群を制御装置と、

(d) 機械翻訳装置モジュールによって翻訳されたテキストを出力しそのテキストを受取人のアドレスに送信するように自動的に選択するための制御入力中の受取人のアドレスの指定に相当する出力装置と、

を有することを特徴とする機械翻訳装置。

30. 請求の範囲21項記載の機械翻訳装置において、制御入力は入力テキストの翻訳に用いられる指定したソース/ター

ネットワークを介して指定した受取人に翻訳した出力テキストを送信する方法において、

(a) 第1通信リンクを介して電子テキストとしての入力テキストを通信リンクネットワーク上のレジスタントユーティリティまたはサーバに送信し、ここで、電子入力テキストはソース言語であり、ターゲット言語に翻訳された出力テキストを受信するように通信リンクネットワークを介してアドレス可能な受取人のアドレスを指定する第1所定フィールドと入力テキストの機械翻訳装置のソース/ターゲット言語対からソース/ターゲット言語対を指定する第2所定フィールドを含み、

(b) ソース言語の入力テキストのターゲット言語の出力テキストへの機械翻訳装置を有する機械翻訳装置モジュールにレジスタントユーティリティまたはサーバでアクセスし、ここで、機械翻訳装置モジュールは指定したソース/ターゲット言語対のサブ言語群を有し、サブ言語群の各々は対応するソース/ターゲット言語対の機械翻訳装置を有するように動作され、機械翻訳装置モジュールは対応するソース/ターゲット言語対のサブ言語群を識別して対応の入力テキストの機械翻訳装置を実行するために制御入力中のソース/ターゲット言語対の指定に相当し、

(c) 入力テキストを作った制御入力に指定された受取人アドレスに通信リンクネットワークを介して通信リンクを介して電子テキストとして指定した出力テキストを自動的に送信する、

ことを特徴とする方法。

31. 請求の範囲21項記載の機械翻訳装置において、制御入力は入力テキストの翻訳に用いられる指定したソース/ター

ネットワークを介して自動的に

順次して順次して送信する方法において、レジデントユーザリ
タイまたはサーバからの入力テキストを隔った時刻に送信機お
よび受信人によって用いられた複数の電子文字符号化基準に定換
する動作をさらに有する方法。」